


| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

УТВЕРЖДЕНО
 решением Ученого совета факультета математики,
 информационных и авиационных технологий
 от «21» июня 2019 г., протокол № 5/19
 Председатель Волков М.А.
(подпись, расшифровка подписи)
 «21» июня 2019 г.



РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

| | |
|------------|---|
| Дисциплина | Методы и системы обработки больших данных |
| Факультет | Факультет математики, информационных и авиационных технологий |
| Кафедра | Информационных технологий |
| Курс | 3 |

Направление (специальность) 09.03.03 «Прикладная информатика»
код направления (специальности), полное наименование

Направленность (профиль/специализация) Информационная сфера
полное наименование

Форма обучения очная
очная, заочная, очно-заочная (указать только те, которые реализуются)

Дата введения в учебный процесс УлГУ: «01» сентября 2019 г.

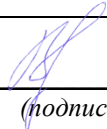
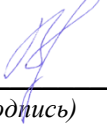
Программа актуализирована на заседании кафедры: протокол № ___ от ___ 20___ г.

Программа актуализирована на заседании кафедры: протокол № ___ от ___ 20___ г.

Программа актуализирована на заседании кафедры: протокол № ___ от ___ 20___ г.

Сведения о разработчиках:

| ФИО | Кафедра | Должность, ученая степень, звание |
|---------------------------|---------------------------|--|
| Волков Максим Анатольевич | Информационных технологий | Заведующий кафедрой, к.ф.-м.н., доцент |

| СОГЛАСОВАНО | СОГЛАСОВАНО |
|---|--|
| Заведующий кафедрой информационных технологий, реализующей дисциплину | Заведующий выпускающей кафедрой информационных технологий |
| /  / <u>Волков М.А.</u> / <i>(подпись) (Ф.И.О.)</i> | /  / <u>Волков М.А.</u> / <i>(подпись) (Ф.И.О.)</i> |
| «21» июня 2019 г. | «21» июня 2019 г. |

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цели освоения дисциплины: освоить принципы, методы, технологии и инструменты использования больших данных в информационных системах в экономике.

Задачи освоения дисциплины: изучить технологии хранения, обработки и анализа больших данных, изучить методы построения информационных систем на основе нереляционных баз данных и распределенных систем хранения.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП

Дисциплина «Методы и системы обработки больших данных» (Б1.В.1.ДВ.03.01) входит в состав вариативной части Основной Профессиональной Образовательной Программы по направлению подготовки 09.03.03 «Прикладная информатика».

Для изучения данной дисциплины необходимы знания основных понятий и методов информатики и программирования, архитектуры вычислительных систем и компьютерных систем, технологии программирования.

Дисциплина закладывает информационные знания необходимые для изучения таких курсов, как параллельное программирование, методы программирования современных информационных систем, объектно-ориентированное программирование.

Результаты освоения дисциплины будут необходимы для дальнейшего процесса обучения в рамках поэтапного формирования компетенций при изучении последующих дисциплин (указаны в ФОС, пункт 1).

3. ПЕРЕЧЕНЬ ПЛАНИРУЕМЫХ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (МОДУЛЮ), СООТНЕСЕННЫХ С ПЛАНИРУЕМЫМИ РЕЗУЛЬТАТАМИ ОСВОЕНИЯ ОПОП

| Код и наименование реализуемой компетенции | Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с индикаторами достижения компетенций |
|--|---|
| ПК-1 способность проводить обследование организаций, выявлять информационные потребности пользователей, формировать требования к информационной системе | Знать: современные приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения и администрирования. Уметь: использовать подобные инструментальные средства в практической деятельности Владеть: применением подобных инструментальных |

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

| | |
|---|--|
| | средств |
| ПК-5 способность моделировать прикладные (бизнес) процессы и предметную область | <p>Знать: современные приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения и администрирования.</p> <p>Уметь: использовать подобные инструментальные средства в практической деятельности</p> <p>Владеть: применением подобных инструментальных средств</p> |
| ПК-7 способность настраивать, эксплуатировать и сопровождать информационные системы и сервисы | <p>Знать: современные приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения и администрирования.</p> <p>Уметь: использовать подобные инструментальные средства в практической деятельности</p> <p>Владеть: применением подобных инструментальных средств</p> |
| ПК-9 способность осуществлять ведение базы данных и поддержку информационного обеспечения решения прикладных задач | <p>Знать: современные приемы работы с инструментальными средствами, поддерживающими создание программных продуктов и программных комплексов, их сопровождения и администрирования.</p> <p>Уметь: использовать подобные инструментальные средства в практической деятельности</p> <p>Владеть: применением подобных инструментальных средств</p> |

4. ОБЩАЯ ТРУДОЕМКОСТЬ ДИСЦИПЛИНЫ

4.1. Объем дисциплины в зачетных единицах (всего) 2

4.2. Объем дисциплины по видам учебной работы (в часах)

| Вид учебной работы | Количество часов (форма обучения <u>очная</u>) | | | |
|---------------------------------|---|-------------------------------|---|---|
| | Всего по плану | В т.ч. по семестрам 6 семестр | | |
| | | 1 | 2 | 3 |
| 1 | 2 | 6 | 7 | 8 |
| Контактная работа обучающихся с | 54 | 54 | | |

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

| | | | | |
|--|-------|-------|--|--|
| преподавателем в соответствии с УП | | | | |
| Аудиторные занятия: | 54 | 54 | | |
| лекции | 18 | 18 | | |
| семинары и практические занятия | | | | |
| лабораторные работы, практикумы | 36 | 36 | | |
| Самостоятельная работа | 18 | 18 | | |
| Форма текущего контроля знаний и контроля самостоятельной работы: тестирование, контр. работа, коллоквиум, реферат и др.(не менее 2 видов) | | | | |
| Курсовая работа | | | | |
| Виды промежуточной аттестации (экзамен, зачет) | зачет | зачет | | |
| Всего часов по дисциплине | 72 | 72 | | |

4.3. Содержание дисциплины (модуля.) Распределение часов по темам и видам учебной работы:

Форма обучения _____ очная _____

| Название разделов и тем | Всего | Виды учебных занятий | | | | | Форма текущего контроля знаний |
|---|-------|----------------------|--------------------------------|---------------------------------|-------------------------------|------------------------|--------------------------------|
| | | Аудиторные занятия | | | Занятия в интерактивной форме | Самостоятельная работа | |
| | | Лекции | Практические занятия, семинары | Лабораторные работы, практикумы | | | |
| 1 | 2 | 3 | 4 | 5 | 6 | 7 | |
| Раздел 1. Основы построения и использования систем больших данных | | | | | | | |
| Тема 1. Основы систем больших данных | 18 | 4 | | 5 | | 9 | |
| Тема 2. Методы | 18 | 5 | | 4 | | 9 | |

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

| | | | | | | | |
|--|----|----|--|----|--|----|--|
| работы с распределенными информационными системами | | | | | | | |
| Раздел 2. Разработка и использование приложений на основе распределенных баз данных | | | | | | | |
| Тема 3. Документно-ориентированные распределенные СУБД | 18 | 5 | | 4 | | 9 | |
| Тема 4. Особенности сбора, хранения, обработки и анализа Больших данных | 18 | 4 | | 5 | | 9 | |
| Итого | 72 | 18 | | 18 | | 36 | |

5. СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Раздел 1. Основы построения и использования систем больших данных

Тема 1. Основы систем больших данных

Понятие Больших данных.

Особенности сбора, хранения, обработки и анализа больших массивов данных.

Источники больших данных.

Использование больших данных в науке, бизнесе, государственном управлении.

Тема 2. Методы работы с распределенными информационными системами

Использование фреймворка

Состав и возможности программного комплекса Apache Hadoop

Раздел 2. Разработка и использование приложений на основе распределенных баз данных

Тема 3. Документно-ориентированные распределенные СУБД

Проектирование структуры документно-ориентированной базы данных.

Понятие агрегата.

Современные документно-ориентированные СУБД.

Тема 4. Особенности сбора, хранения, обработки и анализа Больших данных

Запросы на выборку и модификацию. Использование драйверов.

Настройка фрагментации.

6. ТЕМЫ ПРАКТИЧЕСКИХ И СЕМИНАРСКИХ ЗАНЯТИЙ

Данный вид работы не предусмотрен УП

7. ЛАБОРАТОРНЫЕ РАБОТЫ, ПРАКТИКУМЫ

Лабораторная работа № 1.

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

Знакомство с аналитической платформой Deductor

Цель работы - ознакомиться с архитектурой, основными частями и пользовательским интерфейсом Deductor, получить навыки импорта данных, парциальная предобработка, восстановление пропущенных данных, удаление аномалий, спектральная обработка, удаление шумов.

Содержание работы: Используя текстовый редактор "блокнот" создать файл «TestForPPP.txt», содержащий такие столбцы, как «Аргумент», «Синус», «Аномалии», «Больше шумов», «Средние шумов», «Малые шумов». Разделителем между столбцами является знак табуляции. Столбцу «Аргумент» присваиваются значения от 0 до 2,96 с шагом 0,02. В столбце «Синус» принимаются значения синуса (9 знаков после запятой). При любых двадцати значениях аргумента, ввод данных в значениях синуса пропустить.

Значения столбца «Аномалии» равны значениям столбца «Синус», но не имеют пропущенных данных, однако 10 значений резко отклоняются от истинного значения синуса аргумента. Значения столбцам «Больше шумов», «Средние шумов», «Малые шумов» имеют значения близкие к значению синуса аргумента, но имеют некоторое отклонение (дисперсию и выбираются из промежутка -1,5 до 1,5).

Выполнить импорт данных, созданного файла, обработку данных, восстановить пропущенные значения синуса, выполнить парциальную обработку, удалить аномалии и шумов.

| Аргумент | Синус | Аномалии | Больше шумов | Средние шумов | Малые шумов |
|----------|-------------|-------------|--------------|---------------|-------------|
| 0,14 | 0,139543115 | 0,139543115 | -0,006298541 | 0,126864534 | 0,174892379 |
| 0,16 | 0,159318207 | 0,159318207 | 0,33294777 | 0,085238625 | 0,145575649 |
| 0,18 | 0,179029573 | 0,179029573 | 0,279505602 | 0,085723008 | 0,134508346 |
| 0,2 | 0,198669331 | 0,198669331 | 0,258280061 | 0,277157184 | 0,176827573 |
| 0,22 | 0,218229623 | 0,5 | 0,139744277 | 0,28969863 | 0,231458805 |
| 0,24 | 0,237702626 | 0,237702626 | 0,317646146 | 0,177087062 | 0,282627785 |
| 0,26 | 0,257080552 | 0,257080552 | 0,266913669 | 0,19389017 | 0,288192043 |
| 0,28 | 0,276355649 | 0,276355649 | 0,131932825 | 0,325878695 | 0,243445006 |
| 0,3 | | 0,295520207 | 0,403496144 | 0,30362014 | 0,30123014 |
| 0,32 | | 0,314566561 | 0,430780976 | 0,306273759 | 0,312952364 |

Порядок выполнения работы

Импорт осуществляется путем вызова мастера импорта на панели «Сценарии»

После запуска мастера импорта укажем тип импорта «Текстовый файл с разделителями» и перейдем к настройке импорта. Укажем имя файла, из которого необходимо получить данные (пример для парциальной обработки). В окне просмотра выбранного файла можно увидеть содержание данного файла.

Далее перейдем к настройке параметров импорта. На этой странице мастера предоставляется возможность указать, с какой строки следует начать импорт, указать, то, что первая строка является заголовком, возможность добавить первичный ключ. Указать, что является символом-разделителем столбцов, а также указать ограничитель строк, разделитель целой и дробной части вещественного числа, разделитель компонентов даты и ее формат.

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

В данном случае параметры по умолчанию на этой странице мастера установлены правильно, а именно: начать импорт с первой строки, первая строка является заголовком, разделителем между столбцами является знак табуляции, разделителем целой и дробной частей является запятая. На следующем шаге мастера предоставляется возможность настроить имя, название (метку), размер, тип данных, вид данных и назначение. Некоторые свойства (например, тип данных) можно задавать для выделенного набора столбцов.

Для правильного импорта данных необходимо изменить тип данных у первых трех столбцов («АРГУМЕНТ», «СИНУС», «АНОМАЛИИ»). Тип данных по умолчанию неверный, поскольку программа определяет его, основываясь на значениях первой строки данных. В данном случае там находятся нули - целые числа. Поэтому программа определила, что столбец содержит целочисленные значения. Выделим их с помощью мыши и укажем им тип данных - «Вещественный». Далее осталось только выполнить импорт данных, нажав на кнопку «Пуск» на следующем шаге мастера импорта.

После импорта данных на следующем шаге мастера необходимо выбрать способ отображения данных. В данном случае самым информативным является диаграмма, выберем ее.

От того, какие способы отображения будут выбраны на этом этапе, зависят последующие шаги мастера. В данном случае необходимо настроить, какие столбцы диаграммы следует отображать и как именно. Выберем для отображения поле «СИНУС» и тип диаграммы «Линии».

На последнем шаге мастера необходимо указать название ветки в дереве сценариев. На этом работа мастера импорта заканчивается. Теперь в дереве сценариев появится новый узел с необходимыми данными. В главном окне программы представлены все выбранные отображения данных этого узла. В данном случае только диаграмма.

Далее сделаем обработку данных из файла «TestForPPP.txt». Он содержит таблицу со следующими полями: «АРГУМЕНТ» - аргумент, «СИНУС» - значения синуса аргумента (некоторые значения пустые), «АНОМАЛИИ» - синус с выбросами, «БОЛЬШИЕ ШУМЫ» - значения синуса с большими шумами, «СРЕДНИЕ ШУМЫ» - значения синуса со средними шумами, «МАЛЫЕ ШУМЫ» - значения синуса с малыми шумами. Все данные можно увидеть на диаграмме после импорта из текстового файла.

Импортировав файл можно увидеть, что в столбце «СИНУС» содержатся пустые значения. На диаграмме выше видно, что некоторые значения синуса пропущены. Для дальнейшей обработки необходимо их восстановить. Для этого следует запустить мастер частичной обработки.

Поскольку данные в исходном наборе упорядочены, на следующем шаге мастера обработки выделим поле «СИНУС» и укажем для него тип обработки «Аппроксимация». Так как в данном случае больше ничего не требуется, то остальные параметры обработки оставляем отключенными. Перейдя на страницу запуска процесса обработки, выполняем ее, нажав на пуск, и далее выбираем тип визуализации обработанных данных (как в примере импорта).

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

После выполнения процесса обработки на диаграмме пропуски в данных исчезли, что и было необходимо сделать.

Далее удалим аномалий из поля «АНОМАЛИИ» импортированной таблицы. В мастере парциальной предобработки на третьем шаге выбираем поле «АНОМАЛИИ» и указываем ему тип обработки «Удаления аномальных явлений», степень подавления «Большая».

Так как больше никаких обработок не планировалось, то переходим на шаг запуска процесса обработки и нажимаем «Пуск». После выполнения процесса обработки на диаграмме видно, что выбросы исчезли, остались лишь небольшие возмущения, которые легко сгладить при помощи спектральной обработки.

Аномалии были устранены, однако небольшие возмущения остались. Сгладим их при помощи парциальной обработки. Для этого после удаления аномалий вновь запустим мастер парциальной обработки. В нем на четвертом шаге выберем поле «АНОМАЛИИ» и укажем ему тип обработки «Вейвлет преобразование» с параметрами по умолчанию (глубина разложения 3, порядок вейвлета 6).

Так как больше ничего не планировалось, то перейдем с шагу запуска процесса обработки и выполним ее. В качестве визуализатора укажем диаграмму. После обработки можно убедиться на диаграмме в отсутствии выбросов и сравнить результат с эталонным значением синуса (столбец «СИНУС»).

В примере по парциальной обработке, как было показано ранее, есть 3 столбца с шумами: «БОЛЬШИЕ ШУМЫ», «СРЕДНИЕ ШУМЫ», и «МАЛЫЕ ШУМЫ» - соответственно синус с большими, средними и малыми шумами. Ясно, что для дальнейшей работы с данными эти шумы необходимо устранить. Таким образом, в мастере парциальной обработки на четвертом шаге выберем по очереди поля «БОЛЬШИЕ ШУМЫ», «СРЕДНИЕ ШУМЫ» и «МАЛЫЕ ШУМЫ», зададим тип обработки «Вычитание шума» и укажем степень подавления - «большая», «средняя» и «малая» соответственно. После выполнения обработки на диаграмме можно просмотреть полученные результаты.

Теперь удалим шумы с помощью вейвлет преобразования. В мастере парциальной обработки выберем поля «БОЛЬШИЕ ШУМЫ», «СРЕДНИЕ ШУМЫ» и «МАЛЫЕ ШУМЫ», укажем тип обработки «Вейвлет преобразование», оставив параметры обработки по умолчанию (глубина разложения - 3, порядок вейвлета - 6). На диаграмме можно убедиться в том, что данные сгладились. Повысить качество сглаживания шумов таким способом можно, путем подбора удовлетворительных параметров обработки.

Лабораторная работа №2

Трансформация данных

Цель работы - научиться применять разбиение данных, квантование и фильтрацию для трансформации данных.

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

Создать в блокноте файл «Credit.txt», содержащий данные кредитования. В файле должны быть такие столбцы, как «Сумма кредита», «Дата кредитования» (в формате ДД.ММ.ГГ), «Цель кредитования», «Возраст».

| Сумма кредита | Дата кредитования | Цель кредитования | Возраст |
|---------------|-------------------|----------------------|---------|
| 7000 | 01.01.03 | Оплата услуг | 49 |
| 14578 | 01.01.03 | Покупка товара | 30 |
| 34567 | 03.01.03 | Ремонт недвижимости | 23 |
| 23567 | 04.02.03 | <u>Турпоездка</u> | 22 |
| 7500 | 05.02.03 | Оплата услуг | 56 |
| 12345 | 05.02.03 | Покупка недвижимости | 78 |

2. Импортировать в систему Deductor Studio текстовый файл «Credit.txt».

2.1. Произвести разбиение данных по рискам кредитования физических лиц.

2.2. Получить данные по суммам взятых кредитов по неделям.

2.3. Разбить данные о возрасте кредиторов на 5 интервалов (до 30 лет, от 30 до 40, от 40 до 50, от 50 до 60, старше 60 лет). Причем представить данные в разрезе по неделям.

3. Создать текстовый файл «banks.txt» и импортировать в систему. Файл должен содержать статистику по банкам России за определенный период («Банк», «Филиал», «Город», «Прибыль»)

| Банк | Филиал | Город | Прибыль |
|---------------------|--------|-----------|---------|
| Внешторгбанк | 32 | Москва | 355197 |
| <u>Газпромбанк</u> | 1786 | Казань | 0 |
| АВТОБАНК | 100 | Москва | 4678389 |
| Банк ЗЕНИТ | 24 | Челябинск | 0 |
| "ОАО ""АЛЬФА-БАНК"" | ALFM | Москва | 356564 |

3.1. Выявить ряд городов, в которых прибыль банков самая большая.

Порядок выполнения работы

Интересующие нас столбцы: «СУММА КРЕДИТА», «ДАТА КРЕДИТОВАНИЯ», «ЦЕЛЬ КРЕДИТОВАНИЯ» и «ВОЗРАСТ». После импорта данных из текстового файла наиболее информативно просмотреть данные можно с помощью визуализатора «Куб», выбрав в качестве измерений столбцы «ВОЗРАСТ» и «ЦЕЛЬ КРЕДИТОВАНИЯ», а в качестве факта - столбец «СУММА КРЕДИТА». Остальные столбцы установить как непригодные.

На следующем шаге настройки куба следует указать измерение «ЦЕЛЬ КРЕДИТОВАНИЯ» как измерение в сроках, а измерение «ВОЗРАСТ» как измерение в столбцах, перетащив их с помощью мыши в соответствующие окна из области доступных измерений.

В итоге, на кросс диаграмме (одна из закладок визуализатора куб) можно просмотреть исходные данные.

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

В мастере обработки «Дата и Время» на втором шаге выберем поле «ДАТА КРЕДИТОВАНИЯ» используемым, в появившейся после этого таблице настроек выберем назначение «Используемое» в столбце «Строка» напротив строки «Год + Неделя».

Больше никакие настройки не понадобятся, поэтому перейдем далее к выбору типа визуализации. Выберем в качестве визуализаторов «Таблицу» и «Куб», поставив галочки в соответствующих позициях. В мастере настройки полей куба выберем в качестве измерения появившийся после обработки столбец «ДАТА КРЕДИТОВАНИЯ_YWStr (Год + Неделя)» и столбец «ЦЕЛЬ КРЕДИТОВАНИЯ», а в качестве факта - «СУММА КРЕДИТА». Остальные поля сделаем неиспользуемыми. На следующем шаге перенесем одно измерение из области «доступных» в область «Измерения в строках», а другое - в область «Измерения в столбцах».

Таким образом, на кросс диаграмме имеем суммы взятых кредитов по неделям (за первые две недели года) в разрезе целей кредитования.

В таблице с данными видно, что новое поле - «ДАТА КРЕДИТОВАНИЯ_YWStr (Год + Неделя)» содержит одинаковые значения (дата начала недели) для строк, которые попадают в одну и ту же неделю (дата начала недели или номер недели с начала года).

Далее используем инструмент квантования для разбиения данных о возрасте кредиторов на 5 интервалов (до 30 лет, от 30 до 40, от 40 до 50, от 50 до 60, старше 60 лет). Исходные данные распределятся по пяти интервалам именно так, поскольку, согласно статистике, минимальное значение возраста кредитора 19, а максимальное 69 лет. Это необходимо аналитику для оценки кредиторской активности разных возрастных групп, с целью принятия решения о стимулировании кредиторов в группах с низкой активностью (например, уменьшение стоимости кредита для этих групп) и, быть может, увеличение прибыли в возрастных группах кредиторов с высоким риском (путем предложения дополнительных платных услуг). Причем аналитик желает видеть данные в разрезе по неделям. Воспользуемся мастером квантования.

В нем выберем назначение поля «Возраст» используемым, укажем способ разбиения «По интервалам», зададим количество интервалов равное 5, в качестве значения выберем «Метку интервала». На следующем шаге мастера определим сами метки соответственно возраста кредиторов: «до 30 лет», «от 30 до 40 лет» и т.д.

После обработки выберем в качестве способа отображения «Куб». В мастере укажем «СУММА КРЕДИТА» в качестве факта, «ВОЗРАСТ» и поле «ДАТА КРЕДИТОВАНИЯ (Год + Неделя)» в качестве измерения, остальные поля укажем неиспользуемыми. Далее перенесем «ВОЗРАСТ» из доступных измерений в «Измерения в строках», а «ДАТА КРЕДИТОВАНИЯ (Год + Неделя)» в «Измерения в столбцах». На кросс диаграмме теперь видна информация о том, какие суммы кредитов берут кредиторы определенных возрастных групп в разрезе по неделям.

Теперь аналитик, получив такие данные, может дать рекомендации о снижении стоимости кредита для лиц, старше 50 лет, либо о применении каких -нибудь других мер, способных привлечь большее количество кредиторов этих групп, либо мер, направленных на то, чтобы кредиторы брали кредит на большие суммы.

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

Рассмотрим ситуацию, когда аналитику необходимо спрогнозировать кредитоспособность потенциального кредитора. Предполагается, что кредиторы, берущие суммы разного диапазона ведут себя по-разному, следовательно, модели прогноза должны свои для каждой группы. Т.е. для дальнейшего построения моделей прогноза кредитоспособности определенных аналитиком категорий необходимо использовать фильтрацию.

Определим, для примера группу кредиторов, взявших кредит менее 10000 руб. Воспользуемся данными предыдущего примера. Для этого, находясь на узле импорта данных из текстового файла, запустим мастер обработки. В нем в качестве метода обработки выберем фильтрацию. На втором шаге мастера можно видеть одно неопределенное условие фильтрации (при необходимости их можно добавлять или удалять соответствующими кнопками на форме). Поскольку необходимо отфильтровать данные только по кредиторам, взявших кредит менее 10000, то в графе «Имя поля» выбираем поле «СУММА КРЕДИТА», в графе «Условие» выбираем знак меньше, в графе «Значение» пишем «10000».

Больше никаких условий не требуется, поэтому переходим на следующий шаг мастера и запускаем процесс фильтрации. После выполнения обработки можно манипулировать уже только с данными по кредиторам выбранного кредитного диапазона. В правильности выполненной операции можно легко убедиться, выбрав в качестве визуализации данных статистику и просмотрев значения минимального и максимального значения поля «СУММА КРЕДИТА».

Теперь допустим, что у аналитика имеется статистика по банкам России за определенный период. Перед ним стоит задача выявления ряда городов, в которых прибыль банков самая большая для использования этих данных в дальнейшем. Для этого аналитик должен обратить внимание на следующие поля таблицы из файла: «БАНК», «ФИЛИАЛЫ», «ГОРОД», «ПРИБЫЛЬ». Т.е. информация о названии банка, городе, в котором он находится, (филиалы банка могут находиться в разных городах - следовательно, по одному и тому же банку может быть несколько записей с данными по разным городам) и прибыль банка.

Для решения поставленной задачи первым делом необходимо найти суммарную прибыль всех банков в каждом городе. Для этого и необходима группировка. Для начала следует импортировать данные по банкам из текстового файла. Просмотреть исходную информацию можно в виде куба, где по строкам будут названия банков, а по столбцам - города. С помощью визуализатора «Куб» также можно получить требуемую информацию, выбрав в качестве измерения поле «ГОРОД», а в качестве факта «ПРИБЫЛЬ». Но нам необходимо получить эти данные для последующей обработки, следовательно, необходимо сделать аналогичную группировку.

Находясь в узле импорта, запустим мастер обработки. Выберем в качестве обработки группировку данных. На втором шаге мастера установим назначение поля «ГОРОД» как измерение, а назначение поля «ПРИБЫЛЬ» как факт. В качестве функции агрегации у поля «ПРИБЫЛЬ» следует указать Сумму.

Таким образом, после обработки получим суммарные данные по прибыли всех банков по каждому городу. Их можно просмотреть, используя таблицу.

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

8. ТЕМАТИКА КУРСОВЫХ, КОНТРОЛЬНЫХ РАБОТ, РЕФЕРАТОВ

Данный вид работы не предусмотрен УП

9. ПЕРЕЧЕНЬ ВОПРОСОВ К ЭКЗАМЕНУ (ЗАЧЕТУ)

1. Понятие Больших данных.
2. Особенности сбора, хранения, обработки и анализа Больших данных
3. Требования к распределенным информационным системам
4. Средства построения распределенных информационных систем
5. Технология Map-Reduce
6. Система Apache Hadoop
7. Базы данных NoSQL. Особенности, классификация
8. Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки
9. информации. CAP-теорема.
10. Способы репликации и кластеризации баз данных
11. Документо-ориентированные базы данных
12. Возможности СУБД MongoDB
13. Работа с документо-ориентированными БД на языке JSON.
14. В чем принципиальное отличие концепции Big Data от традиционного подхода BI?
15. Понятие явной (выраженной) и скрытой (структурной) информации.
16. Определение контент-анализа.
17. Каковы основные понятия контент-анализа?
18. Какие существуют виды контент-анализа?
19. Какие существуют этапы контент-анализа?
20. Каковы основные признаки, характеризующие «Большие данные»?
21. Модели развертывания облачных хранилищ.
22. Модели обслуживания облачных хранилищ.
23. Постановка и описание проблемы «последней мили».
24. Безопасность, производительность и надежность при работе с облачными данными.
25. Экономическая составляющая облачных подходов.
26. Способы машинного обучения.
27. Основные фазы обработки «больших данных»

10. САМОСТОЯТЕЛЬНАЯ РАБОТА ОБУЧАЮЩИХСЯ

Форма обучения _____ очная _____

| Название разделов и тем | Вид самостоятельной работы (проработка учебного материала, решение задач, реферат, доклад, контрольная работа, подготовка к сдаче зачета, экзамена и др.) | Объем в часах | Форма контроля (проверка решения задач, реферата и др.) |
|-------------------------|--|---------------|--|
| | | | |

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

| | | | |
|---|---------------------------|---|-------|
| Тема 1. Основы систем больших данных | Подготовка к сдаче зачета | 9 | зачет |
| Тема 2. Методы работы с распределенными информационными системами | Подготовка к сдаче зачета | 9 | зачет |
| Тема 3. Документно-ориентированные распределенные СУБД | Подготовка к сдаче зачета | 9 | зачет |
| Тема 4. Особенности сбора, хранения, обработки и анализа Больших данных | Подготовка к сдаче зачета | 9 | зачет |

11. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

а) Список рекомендуемой литературы

основная

1. Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>
2. Жуковский, О.И. Информационные технологии и анализ данных : учебное пособие / О.И. Жуковский ; Министерство образования и науки Российской Федерации, Томский Государственный Университет Систем Управления и Радиоэлектроники (ТУСУР). - Томск : Эль Контент, 2014. - 130 с. : схем., ил. - Библиогр.: с. 126. - ISBN 978-5-4332-0158-3 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=480500>
3. Базы данных в высокопроизводительных информационных системах : учебное пособие / авт.-сост. Е.И. Николаев ; Министерство образования и науки РФ, Федеральное государственное

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

автономное образовательное учреждение высшего образования «Северо-Кавказский федеральный университет». - Ставрополь : СКФУ, 2016. - 163 с. : ил. - Библиогр.: с.161. ; То же

[Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=466799>

дополнительная

Туманов, В.Е. Проектирование хранилищ данных для систем бизнес-аналитики : учебное пособие / В.Е. Туманов. - Москва : Интернет-Университет Информационных Технологий, 2010.

- 616 с. : ил., табл., схем. - (Основы информационных технологий). - ISBN 978-5-9963-0353-3 ;

То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=233492>

2. Добронец, Б.С. Численный вероятностный анализ неопределенных данных : монография /

Б.С. Добронец, О.А. Попова ; Министерство образования и науки Российской Федерации, Сибирский Федеральный университет. - Красноярск : Сибирский федеральный университет,

2014. - 168 с. : граф., ил. - Библиогр. в кн. - ISBN 978-5-7638-3093-4 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&>

учебно-методическая

1. Рацеев Сергей Михайлович Программирование на языке СИ : учеб. пособие / Рацеев Сергей Михайлович; УлГУ, ФМиИТ. - Ульяновск : УлГУ, 2015. - 319 с. - Библиогр.: с. 317-318.— Режим доступа: <ftp://10.2.96.134/Text/Ratceev2015.pdf>
2. Волков М. А. Методические указания для самостоятельной работы студентов по дисциплине «Методы и системы обработки больших данных» для студентов бакалавриата по направлениям 09.03.03 - «Прикладная информатика», 02.03.03 - «Математическое обеспечение и администрирование информационных систем» / М. А. Волков; УлГУ, ФМИиАТ. - Ульяновск : УлГУ, 2019. - Загл. с экрана; Неопубликованный ресурс. - Электрон. текстовые дан. (1 файл : 249 КБ). - Текст : электронный. <http://lib.ulsu.ru/ProtectedView/Book/ViewBook/10190>

Согласовано:

Главный библиотекарь НБ УлГУ
Должность сотрудника научной библиотеки

/ Полина Н.Ю.
ФИО


подпись

| | | |
|--|-------|---|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

б) Программное обеспечение

Для проведения занятий требуются мультимедийные средства: компьютер с пакетом программ Open Office, Web браузер и проектор.

Для проведения лабораторных работ по курсу «Программирование в среде Windows» требуется компьютерный класс, подключенный к ЛВС УлГУ с выходом в Интернет и с установленным ПО – MS Visual Studio.

в) Профессиональные базы данных, информационно-справочные системы

1. Электронно-библиотечные системы:

1.1. **IPRbooks** [Электронный ресурс]: электронно-библиотечная система / группа компаний Ай Пи Эр Медиа . - Электрон. дан. - Саратов, [2019]. - Режим доступа: <http://www.iprbookshop.ru>.

1.2. **ЮРАЙТ** [Электронный ресурс]: электронно-библиотечная система / ООО Электронное издательство ЮРАЙТ. - Электрон. дан. – Москва, [2019]. - Режим доступа: <https://www.biblio-online.ru>.

1.3. **Консультант студента** [Электронный ресурс]: электронно-библиотечная система / ООО Политехресурс. - Электрон. дан. – Москва, [2019]. - Режим доступа: <http://www.studentlibrary.ru/pages/catalogue.html>.

1.4. **Лань** [Электронный ресурс]: электронно-библиотечная система / ООО ЭБС Лань. - Электрон. дан. – С.-Петербург, [2019]. - Режим доступа: <https://e.lanbook.com>.

1.5. **Znanium.com** [Электронный ресурс]: электронно-библиотечная система / ООО Знаниум. - Электрон. дан. – Москва, [2019]. - Режим доступа: <http://znanium.com>.

2. **КонсультантПлюс** [Электронный ресурс]: справочная правовая система. /Компания «Консультант Плюс» - Электрон. дан. - Москва : КонсультантПлюс, [2019].

3. **База данных периодических изданий** [Электронный ресурс] : электронные журналы / ООО ИВИС. - Электрон. дан. - Москва, [2019]. - Режим доступа: <https://dlib.eastview.com/browse/udb/12>.

4. **Национальная электронная библиотека** [Электронный ресурс]: электронная библиотека. - Электрон. дан. – Москва, [2019]. - Режим доступа: <https://нэб.рф>.

5. **Электронная библиотека диссертаций РГБ** [Электронный ресурс]: электронная библиотека / ФГБУ РГБ. - Электрон. дан. – Москва, [2019]. - Режим доступа: <https://dvs.rsl.ru>.

6. Федеральные информационно-образовательные порталы:

6.1. Информационная система [Единое окно доступа к образовательным ресурсам](http://window.edu.ru). Режим доступа: <http://window.edu.ru>

6.2. Федеральный портал [Российское образование](http://www.edu.ru). Режим доступа: <http://www.edu.ru>

7. Образовательные ресурсы УлГУ:

7.1. Электронная библиотека УлГУ. Режим доступа : <http://lib.ulsu.ru/MegaPro/Web>

7.2. Образовательный портал УлГУ. Режим доступа : <http://edu.ulsu.ru>

| | | |
|--|-------|--|
| Министерство науки и высшего образования РФ Ульяновский государственный университет | Форма |  |
| Ф-Рабочая программа дисциплины | | |

Согласовано:

Зам. начальника УИТиТ
Должность сотрудника УИТиТ
подпись

/ Ключкова А.В.



ФИО

12. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ:

Аудитории для проведения лекций, семинарских занятий, для выполнения лабораторных работ и практикумов, для проведения текущего контроля и промежуточной аттестации, курсового проектирования, групповых и индивидуальных консультаций.

Аудитории укомплектованы специализированной мебелью, учебной доской. Аудитории для проведения лекций оборудованы мультимедийным оборудованием для предоставления информации большой аудитории. Помещения для самостоятельной работы оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа к электронной информационно-образовательной среде, электронно-библиотечной системе.

13. СПЕЦИАЛЬНЫЕ УСЛОВИЯ ДЛЯ ОБУЧАЮЩИХСЯ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) могут предлагаться одни из следующих вариантов восприятия информации с учетом их индивидуальных психофизических особенностей:

– для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации;

– для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

Разработчик



подпись

_____ зав.кафедрой ИТ

_____ М.А.Волков _____

должность

ФИО