Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования
**«Владимирский государственный университет
имени Александра Григорьевича и Николая Григорьевича Столетовых»
(ВлГУ)**

УТВЕРЖДАЮ
Проректор по УМР
_____ А.А.Панфилов
«17» 04 2015 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Интеллектуальней анализ данных
(наименование дисциплины)

Направление подготовки: 01.03.02 «Прикладная математика и информатика»

Профиль/программа подготовки

Уровень высшего образования: бакалавриат

Форма обучения: очная

Семестр	Трудоемкость зач. ед,/ час.	Лекции, час.	Практич. занятия, час.	Лаборат. работы, час.	СРС, час.	Форма промежуточного контроля (экз./зачет)
7	4/144	36	-	18	90	Зачет с оценкой
Итого	4/144	36	-	18	90	Зачет с оценкой

Владимир 2015

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Целью данного курса является знакомство с типами задач, возникающих в области интеллектуального анализа данных (Data Mining) и методами их решения.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО

Дисциплина относиться к обязательным дисциплинам вариативной части ОПОП. Изучение данной дисциплины проходит в 7-м семестре и базируется на знаниях, приобретённых студентами в рамках общеобразовательных и специальных курсов:

- "Теория вероятностей и математическая статистика"
- "Основы информатики"

Для усвоения курса необходимо:

- обладать базовыми знаниями по всем курсам математического и естественно-научного цикла дисциплин, предусмотренных учебном планом
- быть уверенным пользователем основных офисных программ (MS Word, MS Excel)
- уметь самостоятельно работать с различными источниками информации (интернет, печатные издания)

Знания и практические навыки данного курса могут быть применены:

- при написании выпускной квалификационной работы
- в случае продолжения обучения в магистратуре по профильной специальности при изучении курсов, связанных с экспертными системами, искусственным интеллектом, прогнозированием и т.п.
- для профессионального использования при трудоустройстве в IT-компании, занимающиеся автоматизированной обработкой больших массивов данных, разработкой интеллектуальных систем, а так же в любые крупные компании на должности, связанные с аналитической обработкой данных

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)

В результате освоения дисциплины у студента должны быть сформированы общепрофессиональные и профессиональные компетенции, указанные в учебном плане, так же студент должен демонстрировать следующие результаты обучения:

- **Знать:** основные задачи и методы интеллектуального анализа данных (компетенция ПК-2: способность понимать, совершенствовать и применять современный математический аппарат)
- **Владеть:** программными системами анализа данных (компетенция ПК-2: способность понимать, совершенствовать и применять современный математический аппарат; компетенция ОПК-3: способность к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных, тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям)
- **Уметь:** формулировать задачи анализа данных, выбирать адекватные алгоритмы их решения, оценивать качество получаемых решений, а так же самостоятельно с нуля по документации изучать специализированное программное обеспечение (компетенция ОПК-3: способность к разработке алгоритмических и программных решений в области системного и прикладного программирования, математических, информационных и имитационных моделей, созданию информационных ресурсов глобальных сетей, образовательного контента, прикладных баз данных,

тестов и средств тестирования систем и средств на соответствие стандартам и исходным требованиям)

2. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Общая трудоемкость дисциплины составляет 4 зачетные единицы, 144 часов.

№ п/п	Раздел (тема) дисциплины	Семестр	Неделя семестра	Виды учебной работы, включая самостоятельную работу студентов и трудоемкость (в часах)						Объем учебной работы, с применением интерактивных методов (в часах / %)	Формы текущего контроля успеваемости *(по неделям семестра)*, форма промежуточной аттестации *(по семестрам)*
				Лекции	Практические занятия	Лабораторные работы	Контрольные работы	СРС	КП / КР		
1	Введение в дисциплину	7	1	2	-	-	-	-		-	Рейтинг-контроль 1
2	Обзор основных моделей и методов ИАД Обзор типовых бизнес-задач для ИАД по отраслям, обзор некоторых бизнес-задач для ИАД по применению во множестве областей	7	2	2	-	-	-	-	-	1 (50%)	
3	Аналитическая платформа Deductor, как средство для решения основных задач ИАД	7	2	-	-	2	-	30	-	4 (50%)	
			3	2	-	-	-				
			4	2	-	2	-				
4	Анализ качества данных перед их аналитической обработкой, классификация ошибок	7	5	2	-	-	-	-	-	3(50%)	Рейтинг-контроль 2
			6	2	-	2	-				
5	Мониторинг	7	7	2	-	-	-	60	-	9(50%)	

	качества исходных данных на примере анализа данных о продажах в сети розничной торговли		8	2	-	2	-				
			9	2	-	-	-				
			10	2	-	2	-				
			11	2	-	-	-				
			12	2	-	2	-				
6	Поиск ассоциативных правил.	7	13	2	-	-	-	-	-	4(50%)	Рейтинг-контроль № 3
			14	2	-	2	-				
			15	2	-	-	-				
			16	2	-	2	-				
7	Классификация	7	17	2	-	-	-	-	-	3(50%)	
8	Подведение итогов		18	2	-	2	-	-	-	2(50%)	
Всего				36	-	18	-	90		54 (50%)	Зачет с оценкой

Содержание дисциплины

Введение в дисциплину:

- интеллектуальный анализ данных (ИАД) как научное направление;
- понятие Знания, отличие Знаний от Данных, классификация источников Знаний;
- понятие интеллектуального анализа данных (Data Mining), принципиальное отличие от инженерии знаний;
- сферы применения методов ИАД

Обзор основных задач ИАД:

- задача классификации;
- задача регрессии;
- задача прогнозирования;
- задача кластеризации;
- задача поиска ассоциативных правил (задача определения взаимосвязей);
- анализ последовательностей;
- анализ отклонений;
- визуализация

Обзор основных моделей и методов ИАД:

- предсказательные модели;
- описательные модели;
- базовые методы (алгоритмы, основанные на переборе);
- метод деревьев решений (деревья решающих правил, деревья классификации и регрессии);
- нечеткая логика;
- генетические алгоритмы;
- нейронные сети

Обзор типовых бизнес-задач для ИАД по отраслям:

- розничная торговля;
- банковское дело;
- телекоммуникации;
- страхование;
- медицина;
- молекулярная генетика и генная инженерия;
- прикладная химия;
- промышленное производство;
- интернет-технологии

Обзор некоторых бизнес-задач для ИАД по применению во множестве областей:

- прогнозирование при планировании и составлении бюджета;
- маркетинговый анализ;
- анализ работы персонала;
- профилирование клиентов для получения портрета «типичного клиента компании»;
- анализ работы региональных отделений компании;
- сравнительный анализ конкурирующих фирм

Аналитическая платформа Deductor, как средство для решения основных задач ИАД:

- структура платформы;
- возможности экспорта \импорта данных;
- средства очистки и трансформации данных;
- алгоритмы и методы обработки данных;
- средства визуализации

Анализ качества данных перед их аналитической обработкой, классификация ошибок:

- по источнику возникновения;
- по уровню локализации;
- по степени достоверности

Мониторинг качества исходных данных на примере анализа данных о продажах в сети розничной торговли

Поиск ассоциативных правил:

- формальная постановка задачи;
- разновидности задачи поиска ассоциативных правил;
- представление результатов
- алгоритмы

Классификация:

- постановка задачи;
- представление результатов;
- методы построения правил классификации;
- методы построения деревьев решений

Практическая часть курса состоит из трех лабораторных работ и задания, которое выполняется в рамках часов, предусмотренных на самостоятельную работу.

5. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

- лекционно-семинарская система обучения (традиционные лекционные и лабораторные занятия);
- case-study (получение на лабораторных работах учебных кейсов с постановкой задачи и методической проработкой изучаемой темы);

- обучение в малых группах (выполнение лабораторных работ в группах из двух или трёх человек);
- применение мультимедиа технологий (проведение лекционных и семинарских занятий с применением компьютерных презентаций и демонстрационных роликов с помощью проектора или ЭВМ);
- технология развития критического мышления (прививание студентам навыков критической оценки полученных решений)

6. ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ ТЕКУЩЕГО КОНТРОЛЯ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ И УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ

ТЕКУЩИЙ КОНТРОЛЬ

Текущим контролем успеваемости является действующая в университете система рейтинг-контроля. **Контрольным мероприятием рейтинг-конроля являются отчеты по лабораторным работам**. Отчеты сдаются студентами на соответствующих рейтинговых неделях. В зависимости от результатов выполнения лабораторной работы и качества предоставленного отчета студенту выставляется балл рейтинг-контроля.

Лабораторные работы:

1. Знакомство с аналитической платформой Deductor:
   — особенности платной и бесплатной версий;
   — архитектура платформы
   — функционал;
   — разработка отчета
2. Изучение возможностей платформы Deductor для анализа качества исходных данных, а так же решение вопросов, связанных с первичной загрузкой данных на платформу:
   — изучение особенностей внутреннего представления данных на платформе Deductor и поиск путей решения проблемы несовместимости форматов с имеющимися исходными данными к заданию для самостоятельной работы;
   — изучение основных инструментов анализа и возможности их применения для выполнения первого пункта постановки задачи задания для самостоятельной работы
   — разработка отчета;
3. Ассоциативные правила: на базе поиска ассоциативных правил определить типичные шаблоны покупок, совершаемых в супермаркетах (провести анализ покупательской корзины – market basket analysis). Этапы выполнения работы:
   — подготовка данных;
   — формализация задачи и разработка математической модели;
   — поиск в ручном режиме по математической модели типичных шаблонов покупок, оценка найденных шаблонов по критерию достоверности;
   — поиск средствами платформы Deductor по математической модели типичных шаблонов покупок, оценка найденных шаблонов по критерию достоверности;
   — разработка подробного отчета с выводами, в отчете должен быть сравнительный анализ результатов, полученных в ручном режиме и результатов, полученных средствами платформы Deductor

Самостоятельная работа:

Контрольным мероприятием для оценки выполнения студентом самостоятельной работы является итоговый отчет по семестровой работе, который включает реферативную (обзорную) и практическую (аналитическую) составляющие. Отчет разрабатывается поэтапно в течении семестра по мере выполнения семестрового задания к самостоятельной работе. Отчет тщательно проверяется преподавателем, исправляются ошибки, указываются замечания и недочеты, которые студент исправляет. В зависимости от результатов выполнения работы и качества предоставленного отчета студенту выставляется балл, который учитывается в формировании итогового балла промежуточной аттестации.

В рамках самостоятельной работы студенты должны выполнить специальное задание по разделу "Мониторинг качества исходных данных на примере анализа данных о продажах в сети розничной торговли ". Студентам предлагается самостоятельно разбиться на небольшие подгруппы по 2-4 человека, каждая подгруппа получает в качестве исходных данных набор данных от одного филиала сети розничной торговли:

- данные о Филиале;
- сведения о продажах за 1 рабочий день

Данные представлены в виде книг в формате MS Excel:

1. Книга MS Excel Филиал:

- лист Товар, содержит сведения о товарах на складе за 1 месяц (всего 1061 товар);
- лист Группа товаров (всего 26 групп);
- лист Поставщики (всего 16 поставщиков);
- лист Кассир (всего 40 кассиров)

2. Книга MS Excel Продажи:

- лист Продажи, содержит сведения о продажах за 1 рабочий день, из расчета, что время работы всех магазинов розничной сети с 8-00 до 24-00, и в каждом магазине работает 40 кассовых терминалов (всего 30 000 записей)

Исходные данные содержат ошибки:

1. Пропуск в данных
2. Противоречия в записях одного чека
3. Недопустимые значения или сочетания значений
4. "Подозрительные данные" (критерии "подозрительности" основаны на закономерностях, присущих сфере торговли)

Постановка задачи:

1. Подготовить данные о продажах в сети розничной торговли для их последующего анализа в системе Deductor:

- очистка данных, преобразование форматов (формат некоторых данных несовместим с форматами внутреннего представления данных платформы Deductor);
- консолидация данных;
- загрузка данных на платформу Deductor
- мониторинг качества исходных данных: выявление ошибок, их исправление (по возможности);
- выводы о качестве данных и их пригодности для последующего анализа

2. Средствами платформы Deductor провести аналитическую обработку данных по следующим направлениям:

1) анализ потребительской корзины (чеков) для выявления сопутствующих товаров (по группам товаров);

2) анализ динамики продаж:

- выявление суточных пиков продаж для каждой группы товаров;
- выявление суточных пиков "больших" и "малых" чеков по общей сумме чека, распределение сумм покупок в чеках;

- выявление суточных пиков “больших” и “малых” чеков по количеству наименований в одном чеке, определение долей чеков с различным количеством позиций в общем количестве чеков

3. Разработать отчет в формате аналитической записки с подробным описанием выполнения каждого пункта постановки задачи, мотивированными заключениями по каждому пункту исследования и итоговыми выводами. Мотивированные заключения должны сопровождаться наглядными графиками, диаграммами и обобщающими таблицами.

ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ИТОГАМ ОСВОЕНИЯ ДИСЦИПЛИНЫ
(дифференцированный зачет)

Промежуточная аттестация проводится в форме дифференцированного зачета. Итоговая оценка формируется как сумма баллов, набранных студентом в течении семестра. Максимальное количество баллов по всем видам работ в семестре указано в *Таблице 1*

Таблица 1

№	Наименование работ	Максимальное количество баллов по видам работ
1	Рейтинг-контроль 1 - лабораторная работа № 1	15
2	Рейтинг-контроль 2 - лабораторная работа № 2	15
3	Рейтинг-контроль 3 - лабораторная работа № 3	30
4	Выполнение семестрового плана самостоятельной работы – итоговый отчет по семестровой работе	30
5	Посещение занятий	5
6	Дополнительный баллы	5
Итого (максимум)		100

Вопросы к зачету с оценкой

1. Интеллектуальный анализ данных (ИАД) как научное направление;
2. Понятие Знания, отличие Знаний от Данных, классификация источников Знаний;
3. Понятие интеллектуального анализа данных (Data Mining), принципиальное отличие от инженерии знаний;
4. Сферы применения методов ИАД
5. Основные задачи
6. Основные модели и методы ИАД
7. Типовые бизнес-задачи для ИАД по отраслям
8. Аналитическая платформа Deductor, как средство для решения основных задач ИАД
9. Анализ качества данных перед их аналитической обработкой, классификация ошибок
10. Ассоциативные правила
11. Классификация

7. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

а) основная литература:

1. Федин Ф.О. Анализ данных. Часть 1. Подготовка данных к анализу [Электронный ресурс]: учебное пособие/ Федин Ф.О., Федин Ф.Ф.— Электрон. текстовые данные.— М.: Московский городской педагогический университет, 2012.— 204 с.— Режим доступа: http://www.iprbookshop.ru/26444.— ЭБС «IPRbooks», по паролю
2. Федин Ф.О. Анализ данных. Часть 2. Инструменты Data Mining [Электронный ресурс]: учебное пособие/ Федин Ф.О., Федин Ф.Ф.— Электрон. текстовые данные.— М.: Мос-

ковский городской педагогический университет, 2012.— 308 с.— Режим доступа: http://www.iprbookshop.ru/26445.— ЭБС «IPRbooks», по паролю

3. Интеллектуальные системы [Электронный ресурс]: учебное пособие/ А.М. Семенов [и др.].— Электрон. текстовые данные.— Оренбург: Оренбургский государственный университет, ЭБС АСВ, 2013.— 236 с.— Режим доступа: http://www.iprbookshop.ru/30055.— ЭБС «IPRbooks», по паролю
4. Моделирование систем. Подходы и методы [Электронный ресурс]: учебное пособие/ В.Н. Волкова [и др.].— Электрон. текстовые данные.— СПб.: Санкт-Петербургский политехнический университет Петра Великого, 2013.— 568 с.— Режим доступа: http://www.iprbookshop.ru/43957.— ЭБС «IPRbooks», по паролю

б) дополнительная литература:

1. Воловиков С.А. Экономические прогнозы по временным рядам [Электронный ресурс]: учебное пособие/ Воловиков С.А.— Электрон. текстовые данные.— М.: Московский городской педагогический университет, 2010.— 34 с.— Режим доступа: http://www.iprbookshop.ru/26665.— ЭБС «IPRbooks», по паролю

в) интернет-ресурсы:

1) *Арустамов А.* Анализ бизнес информации - основные принципы. Режим доступа: http://www.basegroup.ru/library/methodology/analysisbusinessdata/
2) *Арустамов А.* Анализ больших объемов данных. Режим доступа: http://www.basegroup.ru/library/methodology/very_large_data/
3) *Паршина А., Арустамов А., BaseGroup Labs.* Применение Data Mining для повышения лояльности клиентов. Режим доступа: http://www.basegroup.ru/library/practice/data_mining_in_loyality
4) *Стариков А.* Сегментация данных как метод сравнительного анализа. Режим доступа: http://www.basegroup.ru/library/practice/comparative_analysis
5) *Паклин Н.* Применение логистической регрессии в медицине и скоринге. Режим доступа: http://www.basegroup.ru/library/practice/logis_medic_scoring
6) *Морозов А.* Оценка эффективности рекламной кампании. Режим доступа: http://www.basegroup.ru/library/practice/promotion
7) *Арустамов А.* Разбор адреса на составляющие. Режим доступа: http://www.basegroup.ru/library/cleaning/addresses/
8) *Шахиди А.* Apriori - масштабируемый алгоритм поиска ассоциативных правил. Режим доступа: http://www.basegroup.ru/library/analysis/association_rules/apriori/
9) *Ларин С.* Выявление обобщенных ассоциативных правил.. Режим доступа: http://www.basegroup.ru/library/analysis/association_rules/generalized/
10) *Орешков В.* FPG - альтернативный алгоритм поиска ассоциативных правил. Режим доступа: http://www.basegroup.ru/library/analysis/association_rules/fpg/
11) *Орешков В.* Поиск последовательных шаблонов. Часть 1. Режим доступа: http://www.basegroup.ru/library/analysis/association_rules/sequential_patterns_1/
12) *Орешков В.* Поиск последовательных шаблонов. Часть 2. Режим доступа: http://www.basegroup.ru/library/analysis/association_rules/sequential_patterns_2/
13) К. Канаян, Р. Канаян. Инструменты розничного аналитика. Режим доступа: http://www.shop-academy.com/articles/assortment-policy/retail-analysis-059.htm

г) программное обеспечение (ПО):

- платформа для создания законченных аналитических решений Deductor 5.3. Academic. Режим для скачивания и установки: https://basegroup.ru/sites/default/files/deductor_academic_5.3.0.88.zip

ПО – российское, версия Academic является бесплатной, находится в открытом доступе и загружается для последующей установки с официального сайта компании-разработчика BaseGroup Labs.

3. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Лекционные аудитории, оснащённые доской (для мела или маркера), экраном для проекционных систем, проектором и ноутбуком (420-3, 430-3).

Аудитории для проведения лабораторных занятий, оснащённые современными персональными компьютерами, объединёнными в локальную вычислительную сеть и укомплектованными необходимым системным и прикладным программным обеспечением (122б-3, 100-3, 511-3), аудитории вычислительного центра.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению 01.03.02 Прикладная математики и информатика.

Рабочую программу составил ст. препод. кафедры ФиПМ Воронова Н.М. _____

Рецензент

(представитель работодателя) генеральный директор ООО "РС Сервис" Иванов Д.С.

(место работы, должность, ФИО, подпись)

Программа рассмотрена и одобрена на заседании кафедры ФиПМ

Протокол № 11А от 17.04.15 года

Заведующий кафедрой _____ С.М. Аракелян

(ФИО, подпись)

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии направления 01.03.02 Прикладная математика и информатика

Протокол № 11А от 17.04.15 года

Председатель комиссии _____ С.М. Аракелян

(ФИО, подпись)

ЛИСТ ПЕРЕУТВЕРЖДЕНИЯ
РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ (МОДУЛЯ)

Рабочая программа одобрена на 18-19 учебный год

Протокол заседания кафедры № 1 от 03.09.18 года

Заведующий кафедрой _____ С.М. Аракелян

Рабочая программа одобрена на 19-20 учебный год

Протокол заседания кафедры № 1 от 02.09.19 года

Заведующий кафедрой _____ С.М. Аракелян

Рабочая программа одобрена на 2020-2021 учебный год

Протокол заседания кафедры № 1 от 31.08.2020 года

Заведующий кафедрой _____ С.М. Аракелян