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

УТВЕРЖДАЮ
Проректор
по образовательной деятельности
_____А.А. Панфилов
«30» августа 2016 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
Анализ данных
(наименование дисциплины)

Направление подготовки 38.03.05 «Бизнес-информатика»

Профиль/программа подготовки Бизнес-информатика

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	3/108	18	36		54	Зачет
Итого	**3/108**	**18**	**36**		**54**	**Зачет**

Владимир, 2016

1. Цели освоения дисциплины

Цель дисциплины: формирование у студентов теоретических знаний, практических умений и навыков по применению современных методов аналитической обработки электронных массивов данных в различных сферах человеческой деятельности.

Задачи дисциплины:

□ изучение существующих технологий подготовки данных к анализу;

□ изучение основных методов поиска в данных внутренних закономерностей, взаимосвязей, тенденций;

□ овладение практическими умениями и навыками реализации технологий аналитической обработки данных, формирования и проверки гипотез о их природе и структуре, варьирования применяемыми моделями;

□ формирование умений и навыков применения универсальных программных пакетов и аналитических платформ для анализа данных.

2. Место дисциплины в структуре ОПОП ВО

Место дисциплины в учебном процессе: дисциплина «Анализ данных» относится к базовой части подготовки бакалавра. Учебная дисциплина «Анализ данных» базируется на изучении дисциплин «Теория вероятностей и математическая статистика», «Имитационное моделирование», «Хранилища данных».

3. Компетенции обучающегося, формируемые в результате освоения дисциплины

У обучающегося должны быть сформированы следующие профессиональные компетенции: Способность использовать соответствующий математический аппарат и инструментальные средства для обработки, анализа и систематизации информации по теме исследования (ПК-18)

В результате изучения дисциплины студент должен:

знать:

□ проблемы и направления научных изысканий в области аналитической обработки данных;

□ тенденции разработки универсальных программных средств и аналитических платформ, предназначенных для аналитической обработки данных, построения прогнозов и аналитических сценариев;

□ основные методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных;

□ принципы построения и структурную организацию хранилищ данных;

□ алгоритмы поиска ассоциативных правил и кластерного анализа;

□ статистические и машинные методы классификации и регрессии;

□ методики анализа и прогнозирования временных рядов;

□ технологию построения ансамблей и сравнения моделей;

□ возможности отечественных и зарубежных универсальных программных средств и аналитических платформ, применяемых для анализа данных;

□ проблемные вопросы внедрения аналитических программных продуктов и технологий в профессиональную деятельность организаций и учреждений;

уметь:

□ практически применять методы консолидации, трансформации, визуализации, оценки качества, очистки и предобработки данных для качественной подготовки данных к анализу;
□ применять технологии анализа электронных массивов данных для решения конкретных практических проблем;
□ использовать возможности отечественных и зарубежных универсальных программных средств и аналитических платформ для аналитической обработки данных, построения прогнозов и аналитических сценариев;
□ свободно ориентироваться на современном динамичном рынке аналитических программных продуктов.

Владеть:
□ технологиями аналитической обработки электронных массивов данных в целях решения практических проблем выбранной предметной области;
□ навыками выбора и применения отечественных и зарубежных аналитических платформ, используемых для анализа табулированных массивов электронных данных.

4. Структура и содержание дисциплины «Анализ данных»

Общая трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

№ п/п	Раздел 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 Форма промежуточной аттестации
				лекции	Лаб. работа	Практическая работа.	Самост. работа	КР		
1.	**Тема 1**. Аффинитивный анализ. Поиск последовательных шаблонов. Введение в аффинитивный анализ (affinity analysis). Алгоритм a priori. Иерархические ассоциативные правила.	7	1-2	2		4	6		3/50%	РК 1
2.	**Тема 2.** Кластерный анализ. Введение в кластеризацию. Классификация методов кластеризации. Алгоритм		3-5	4		4	6		4/50%	

	кластеризации k-means. Сети Кохонена (KCN – Kohonen network). Карты Кохонена (SOM – self organizing map). Проблемы алгоритмов кластеризации.									
3.	**Тема 3.** Классификация и регрессия. Статистические методы. Введение в классификацию и регрессию. Простая линейная регрессия. Оценка соответствия простой линейной регрессии реальным данным. Простая регрессионная модель. Множественная линейная регрессия. Модель множественной линейной регрессии. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Основы логистической регрессии. Интерпретация модели логистической регрессии.		6-8	4		8	12		5/50%	РК-2
4.	**Тема 4.** Классификация и регрессия. Машинное обучение. Введение в деревья решений. Алгоритмы построения деревьев решений. Алгоритмы ID3 и C4.5. Алгоритм CART. Упрощение		9-12	4		4	9		4/50%	

	деревьев решений. Введение в нейронные сети. Искусственный нейрон. Принципы построения нейронных сетей. Алгоритмы обучения нейронных сетей. Алгоритм обратного распространения ошибки.									
5.	**Тема 5.** Анализ и прогнозирование временных рядов. Временной ряд и его компоненты. Модели прогнозирования. Прогнозирование в торговле и логистике.		12-15	2		8	9		5/50%	РК-3
6.	**Тема 6.** Ансамбли моделей. Введение в ансамбли моделей. Бэггинг. Бустинг. Альтернативные методы построения ансамблей.		16-18	2		8	12		5/50%	
	Итого			18		36	54		27/50%	Зачет

СОДЕРЖАНИЕ КУРСА

4.2. Теоретическая часть

Тематика лекций.

Тема 1. Аффинитивный анализ. Поиск последовательных шаблонов

Тема 2. Кластерный анализ.

Тема 3. Классификация и регрессия. Статистические методы.

Тема 4. Классификация и регрессия. Машинное обучение

Тема 5. Анализ и прогнозирование временных рядов.

Тема 6. Ансамбли моделей.

4.3. Практические занятия

Практические занятия являются одним из важнейших видов учебных занятий. Они способствуют максимально эффективному закреплению изучаемого материала на основе углубленной самостоятельной работы студентов в процессе подготовки к занятиям, а также активного участия в самих занятиях.

На занятиях студенты, опираясь на изученные материалы, выступают с индивидуальными докладами, участвуют в общей дискуссии, сопоставляя самые разнообразные мнения и суждения специалистов, высказывают собственные взгляды по наиболее важным, существенным по-

ложениям соответствующих разделов изучаемого курса. В соответствии с проблематикой возможны самые различные виды семинарских занятий. Однако, в конечном итоге, каждое из них предполагает постоянную активность всех участников дискуссии, аналитический подход к изучаемому материалу, выделение в нем главного, основного, формулирование выводов и т.д.

При подготовке к семинару рекомендуется вести конспект, специально отмечая в нем те положения, с которыми в первую очередь намерен выступить участник общей дискуссии или докладчик. Во время семинара важна научно и политически корректная постановка обсуждаемых вопросов, сопровождаемая высокой активностью участников дискуссии. Итоги практических и семинарских занятий учитываются при выставлении общей оценки по предмету.

Темы практических занятий:
Практическое занятие № 1. Простейшие вычисления в MATLAB
Практическое занятие № 2. Работа с массивами в MATLAB
Практическое занятие № 3. Основы программирования в MATLAB
Практическое занятие № 4. Исследование функций на основе численных методов в MATLAB
Практическое занятие № 5. Создание пользовательских интерфейсов в MATLAB
Практическое занятие № 6. Аппроксимация и интерполяция данных в MATLAB
Практическое занятие № 7. Моделирование в приложении Simulink

4.4. Самостоятельная работа студентов

Целью самостоятельной работы являются формирование личности студента, развитие его способности к самообучению и повышению своего профессионального уровня.

Основные формы самостоятельной работы заключаются в проработке дополнительной литературы, подготовке к практическим занятиям, устному опросу, контрольным работам и рейтинг-контролю. Контроль за самостоятельной работой студентов осуществляется на консультациях, во время работы на ПК и практических занятиях.

5. Образовательные технологии

1. лекционно-семинарская система обучения (традиционные лекционные и практические занятия);
2. обучение в малых группах (выполнение практических работ в группах из двух или трёх человек);
3. применение мультимедиа технологий (проведение лекционных и практических занятий с применением компьютерных презентаций и демонстрационных роликов с помощью проектора или ЭВМ);
4. технология развития критического мышления (прививание студентам навыков критической оценки предлагаемых решений);
5. информационно-коммуникационные технологии (применение информационных технологий для мониторинга текущей успеваемости студентов и контроля знаний);

Объем занятий, проводимых в интерактивной форме, составляет 50%.

6. Оценочные средства для текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины и учебно-методическое обеспечение самостоятельной работы студентов

В рамках документа «Положение о рейтинговой системе комплексной оценки знаний студентов» разработан регламент проведения и оценивания контрольных действий. Процедура оценивания знаний, умений, навыков по дисциплине включает учёт успешности выполнения ряда мероприятий: текущего контроля (контрольных работ, рейтинг – контролей); самостоятельной работы (типовых расчетов, курсовых работ и др.) и промежуточной аттестации (зачёта, зачета с оценкой или экзамена).

Публикуемые компоненты ФОС:

1. Полный список теоретических вопросов промежуточной аттестации (несменяемая часть).
2. Типовые формы текущей аттестации (список тем).
3. Список вопросов к самостоятельной работе.

Для генерирования сменяемой части оценочных средств (задач), используются материалы библиотеки ВлГУ и указанных там же специальных сайтов.

Текущий контроль в форме рейтинг -контролей.

Рейтинг-контроль 1

1. Дисперсионный анализ данных.
2. Кластерный анализ данных.
3. Методы классификации в Data mining.
4. Регрессионный анализ данных.
5. Анализ данных с использованием бинарной логистической регрессии.
6. Факторный анализ данных.
7. Метрики, применяемые в Data mining.
8. Ковариационный анализ данных.
9. Методы поиска ассоциативных правил.
10. Сиквенциальный анализ даных (поиск последовательных шаблонов).
11. Основные стандарты Data mining.
12. Анализ данных с использованием сети Кохонена.

Рейтинг-контроль 2

1.Характеристики инструментальных средств Data mining.
2. Реляционные хранилища данных.
3. Многомерные хранилища данных.
4. Гибридные хранилища данных.
5. Виртуальные хранилища данных.
6. Характеристика ETL-процесса.
7. Оценка качества, очистка и предобработка анализируемых данных.
8. Сокращение размерности исходного множества анализируемых данных.
9. Искусственные нейронные сети. Многослойный персептрон.
10. Анализ данных с использованием генетических алгоритмов.
11. Анализ данных с использованием самоорганизующихся карт.
12. Оценка значимости регрессионных моделей с применением t-критерия Стьюдента

Рейтинг-контроль 3

1. Оценка значимости регрессионных моделей с применением F-критерия Фишера.
2. Алгоритм построения деревьев решений ID3

3. Алгоритм построения деревьев решений C4.5.
4. Оценка полезности, эффективности и точности моделей, применяемых для анализа данных.
5. Анализ данных с использованием ансамблей моделей.
6. Проблемы обучения и переобучения моделей.
7. Технологии обогащения данных.
8. Повышение эффективности моделей с помощью бэггинга и бустинга.
9.Градиентный алгоритм обучения многослойного персептрона (алгоритм обратного распространения ошибки).
10. Lift и Profit-кривые.
11. ROC-анализ.

Самостоятельная работа студентов

Список вопросов

1. Практическое применение алгоритмов Data mining.
2. Классификация с несколькими независимыми переменными методом Naïve Bayes.
3. Поиск оптимальной функции методом наименьших квадратов.
4. Сиквенциальный анализ.
5. Меры близости, основанные на расстояниях, используемые в алгоритмах кластеризации.
6. Кластеризация данных при помощи нечетких отношений.
7. Характеристика классов задач, решаемых методами Data Mining.
8. Стандарты Data mining. Характеристика стандартов CWM и PMLL.
9. Библиотеки доступа к алгоритмам Data mining. Характеристика библиотеки Xelopes.
10. Характеристика программных инструментов для выполнения интеллектуального анализа данных.
11. Общая характеристика и классификация методов кластерного анализа данных.
12. Анализ данных с использованием методов классификации и регрессии.
13. Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.
14. Сферы применения деревьев решений.
15. Цели, задачи и принципы работы нейронных сетей.
16. Алгоритмы обучения нейронных сетей.
17. Цели, задачи, принципы и модели прогнозирования.

Промежуточная аттестация в форме зачета

Вопросы к зачету

1. Модели и их свойства. Аналитический и информационный походы к моделированию.
2. Формы представления, типы и виды анализируемых данных.
3. Обучение моделей «с учителем» и «без учителя». Обучающее и тестовое множество. Ошибки обучения. Эффект переобучения.
4. Общая схема анализа данных. Требования к алгоритмам анализа данных.
5. Характеристика этапов технологии KDD.
6. Data Mining. Характеристика классов задач, решаемых методами Data Mining.
7. Программный инструментарий для выполнения анализа данных.
8. Основные положения концепции хранилищ данных (DW).

9. Цели и задачи аффинитивного анализа. Поддержка и достоверность ассоциативных правил. Лифт и левередж.
10. Сферы применения ассоциативных правил.
11. Иерархические ассоциативные правила.
12. Цели, задачи и основное содержание кластерного анализа. Классификация методов кластеризации.
13. Способы определения меры расстояния между кластерами.
14. Характеристика методов связи для процедуры кластеризации (одиночная, полная, средняя).
15. Алгоритм кластеризации k-means.
16. Сети Кохонена (KCN).
17. Карты Кохонена (SOM).
18. Проблемы алгоритмов кластеризации.
19. Цели, задачи и отличительные особенности классификации и регрессии.
20. Сферы применения методов классификации и регрессии.
21. Простая линейная регрессия.
22. Оценка соответствия простой линейной регрессии реальным данным.
23. Простая регрессионная модель.
24. Оценка значимости простой регрессионной модели (t-критерий и F-критерий).
25. Множественная линейная регрессия.
26. Модель множественной линейной регрессии.
28. Оценка значимости множественной регрессионной модели.
29. Регрессия с категориальными входными переменными.
30. Методы отбора переменных в регрессионные модели.
31. Ограничения применимости регрессионных моделей.
32. Логистическая регрессия. Интерпретация модели логистической регрессии.
33. Множественная логистическая регрессия.
34. Цели, задачи и принципы построения деревьев решений. Общая характеристика алгоритмов построения деревьев решений.
35. Сферы применения деревьев решений.
36. Алгоритмы ID3 и C4.5.
37. Алгоритм CART.
38. Упрощение деревьев решений.
39. Цели, задачи и принципы работы нейронных сетей.
40. Принципы функционирования многослойного персептрона.
41. Алгоритмы обучения нейронных сетей.
42. Алгоритм обратного распространения ошибки.
43. Общая характеристика временных рядов и их компонентов. Цели и задачи анализа временных рядов.
44. Цели, задачи и принципы прогнозирования. Модели прогнозирования. Обобщенная модель прогноза.
45. Ансамбли моделей. Бэггинг. Бустинг.
46. Альтернативные методы построения ансамблей.
47. Оценка эффективности и сравнение моделей.
48. Lift- и Profit-кривые.
49. ROC-анализ.

7. Учебно –методическое и информационное обеспечение дисциплины

Основная литература

1. Прикладные методы анализа статистических данных [Электронный ресурс]: учеб. пособие / Горяинова Е.Р., Панков А.Р., Платонов Е.Н. - М. : ИД Высшей школы экономики, 2012. -- 310, [2] с. - 1000 экз. - ISBN 978-5-7598-0866-4.
http://www.studentlibrary.ru/book/ISBN9785759808664.html
2. Эконометрика [Электронный ресурс] : учебник / под ред. д-ра экон. наук, проф. В.С. Мхитаряна. - М. : Проспект, 2014. - 384 с. - ISBN 978-5-392-13469-4.
http://www.studentlibrary.ru/book/ISBN9785392134694.html
3. Популярное введение в современный анализ данных в системе STATISTICA [Электронный ресурс] : Учебное пособие для вузов / Боровиков В.П. - М. : Горячая линия - Телеком, 2013. - 288 с., ил. - ISBN 978-5-9912-0326-5. http://www.studentlibrary.ru/book/ISBN9785991203265.html

Дополнительная литература

1. Информационные системы в экономике [Электронный ресурс] / Горбенко А.О. - М. : БИНОМ, 2013, - 292 с.: ил. - ISBN 978-5-9963-2268-8.
http://www.studentlibrary.ru/book/ISBN9785996322688.html
2. Введение в эконометрику. [Электронный ресурс] / Артамонов Н.В. - М.: МЦНМО, 2011. - 204 с. - ISBN 978-5-94057-727-0. http://www.studentlibrary.ru/book/ISBN9785940577270.html
3. Исследование операций для экономистов, политологов и менеджеров [Электронный ресурс] / Токарев В.В. - М. : ФИЗМАТЛИТ, 2014,- 408 с. - ISBN 978-5-9221-1451-6.
http://www.studentlibrary.ru/book/ISBN9785922114516.html

Периодическая литература (журналы)

1. Информационные технологии.
2. Математическое моделирование
3. Программная инженерия

8. Материально-техническое обеспечение дисциплины

1.Для проведения лекционных занятий: требуется аудитория, оборудованная меловой доской, интерактивной доской, мультимедийным проектором с экраном.
2.Для проведения лабораторных работ: требуется специализированный компьютерный класс.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению 38.03.05 «Бизнес-информатика»

Рабочую программу составила д.т.н., профессор А.А. Малафеева _____
(ФИО, подпись)

Рецензент
(представитель работодателя) директор по маркетингу ЗАО Инвестиционная фирма «ПРОК - Инвест» _____О.В. Крисько
(место работы, должность, ФИО, подпись)

Программа рассмотрена и одобрена на заседании кафедры ФАиП

Протокол № 10 от 30.08.2016 года

Заведующий кафедрой _____А.А. Давыдов
(ФИО, подпись)

Рабочая программа рассмотрена и одобрена на заседании учебно-методической комиссии направления 38.03.05 "Бизнес-информатика"

Протокол № 1 от 30.08.2016 года

Председатель комиссии _____Н.Б. Гисченко
(ФИО, подпись)

ЛИСТ ПЕРЕУТВЕРЖДЕНИЯ
РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ (МОДУЛЯ)

Рабочая программа одобрена на _____ учебный год

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой_____

Рабочая программа одобрена на _____ учебный год

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой_____

Рабочая программа одобрена на _____ учебный год

Протокол заседания кафедры № _____ от _____ года

Заведующий кафедрой_____