

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение
высшего профессионального образования

«УФИМСКИЙ ГОСУДАРСТВЕННЫЙ АВИАЦИОННЫЙ
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»

Кафедра вычислительной математики и кибернетики

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ

УЧЕБНОЙ ДИСЦИПЛИНЫ

«АНАЛИЗ СТАТИСТИЧЕСКИХ ДАННЫХ»

Уровень подготовки: высшее образование – подготовка магистров

Направление подготовки магистров

38.04.05 Бизнес-информатика

(код и наименование направления подготовки)

Направленность подготовки

Бизнес-аналитика

(наименование программы подготовки)

Квалификация (степень) выпускника

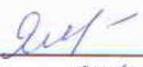
Магистр.

Форма обучения

очная

2015

Исполнители:

к.ф.-м. н. доцент		Ямилова Л.С.
должность	подпись	расшифровка подписи
к.т.н. доцент		Лакман И.А.
должность	подпись	расшифровка подписи
Заведующий кафедрой ВМиК		Юсупова Н.И.

Место дисциплины в структуре образовательной программы

Дисциплина Анализ статистических данных является дисциплиной по выбору *вариативной* части ОПОП по направлению подготовки **38.04.05 Бизнес-информатика**, направленность: **Бизнес-аналитика**.

Рабочая программа составлена в соответствии с требованиями Федерального государственного образовательного стандарта высшего образования по направлению подготовки магистров 38.04.05 Бизнес-информатика, утвержденного приказом Министерства образования и науки Российской Федерации от "08" апреля 2015 г. № 370. Является неотъемлемой частью основной образовательной профессиональной программы (ОПОП).

Целью освоения дисциплины является формирование у будущих магистров в области бизнес-аналитики теоретических знаний и практических навыков для решения научно-исследовательских и прикладных задач связанных со статистическим анализом данных и принятием решений на основе проведенного анализа.

Задачи:

- обучение магистрантов комплексному статистическому анализу данных для решения задач аналитических исследований в области IT.;
- Научить магистранта применять методы дискриминантного анализа и байесовского классификатора для решения задач распознавания.
- Приучить проверять аналитические исследования на качество.
- Научить решать задачи классификации.
- Привить навыки обязательной проверки качества решаемых задач классификации с помощью стандартных статистических критериев.
- Сформировать опыт применения современных статистических пакетов прикладных программ для проведения анализа статистических данных.

Входные компетенции:

На пороговом уровне ряд компетенций был сформирован за счет обучения на предыдущих уровнях высшего образования (специалитет, бакалавриат).

№	Компетенция	Код	Уровень освоения, определяемый этапом формирования компетенции*	Название дисциплины (модуля), практики, научных исследований, сформировавших данную компетенцию
1	Способность применять методы системного анализа и моделирования для анализа, архитектуры предприятий	ПК-3	базовый уровень первого этапа освоения компетенции (параллельно)	Системный анализ

*- **пороговый уровень** дает общее представление о виде деятельности, основных закономерностях функционирования объектов профессиональной деятельности, методов и алгоритмов решения практических задач;

-**базовый уровень** позволяет решать типовые задачи, принимать профессиональные и управленческие решения по известным алгоритмам, правилам и методикам;

-**повышенный уровень** предполагает готовность решать практические задачи повышенной сложности, нетиповые задачи, принимать профессиональные и управленческие решения в условиях неполной определенности, при недостаточном документальном, нормативном и методическом обеспечении.

Исходящие компетенции:

№	Компетенция	Код	Уровень освоения, определяемый этапом формирования компетенции	Название дисциплины (модуля), практики, научных исследований для которых данная компетенция является входной
1	Способность применять методы системного анализа и моделирования для анализа, архитектуры предприятий	ПК-3	Продвинутый уровень, второй этап	Имитационное моделирование
2	Способность применять методы системного анализа и моделирования для анализа, архитектуры предприятий	ПК-3	Продвинутый уровень, третий этап	Государственная итоговая аттестация
3	Готовность к саморазвитию, самореализации, использованию творческого потенциала	ОК-3	Продвинутый уровень, второй этап	Научно-исследовательская работа

Перечень результатов обучения

Процесс изучения дисциплины направлен на формирование элементов следующих компетенций. Планируемые результаты обучения по дисциплине

№	Формируемые компетенции	Код	Знать	Уметь	Владеть
1	Способность готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ	ПК-1	- методы анализа мощности; -методы и подходы проведения дискриминантного анализа; - методы проведения анализа соответствий;	- применять методы оценивания величины экспериментальных эффектов на основе анализа мощности; - рассчитывать различные критерии для определения «нагруженности» дискриминантных функций; -представлять матрицы входной информации в различных видах для удобства проведения анализа соответствий	- опытом подготовки аналитических отчетов на основе проведения анализа соответствий; - опытом подготовки отчетов по классификации и выявлению самых значимых факторов , определяющих классификацию на основе проведения дискриминантного анализа. -опытом проведения анализа мощности для критериев при принятии решений.
2	Способность консультировать по вопросам развития ИТ-инфраструктуры предприятия	ПК-15	- подходы к решению задач классификации ИТ-приложений на основе методов байесовского классификатора.	проводить анализ на основе наивного байесовского классификатора для решения задач классификации	- навыками решению задач классификации ИТ-приложений на основе методов байесовского классификатора

Содержание и структура дисциплины (модуля)

Общая трудоемкость дисциплины составляет 4 зачетные единицы (144 часа).

Трудоемкость дисциплины по видам работ

Вид работы	Трудоемкость, час.	
	1 семестр 144 часов /4 ЗЕ	
Лекции (Л)	12	
Практические занятия (ПЗ)	4	
Лабораторные работы (ЛР)	16	
КСР	3	
Курсовая проект работа (КР)		
Расчетно-графическая работа (РГР)		
Самостоятельная работа (проработка и повторение лекционного материала и материала учебников и учебных пособий, подготовка к лабораторным и практическим занятиям, коллоквиумам, рубежному контролю и т.д.)	100	
Подготовка и сдача экзамена		
Подготовка и сдача зачета (контроль)	9	
Вид итогового контроля (зачет, экзамен)	зачет	

Содержание разделов и формы текущего контроля

№	Наименование и содержание раздела	Количество часов					Литература, рекомендуемая студентам*	Виды интерактивных образовательных технологий**	
		Аудиторная работа				СРС			Всего
		Л	ПЗ	ЛР	КСР				
1	Анализ мощности выборки: Методы формирования выборок и определение их необходимого объема в зависимости от цели проводимого исследования. Простая случайная, стратификационная и серийная выборки. Анализ мощности статистических критериев. Критерий значимости (RS/AS). Углубленные методы доверительного интервального оценивания. Графические и аналитические процедуры анализа мощности. Углубленные методы оценивания величины экспериментальных эффектов. Нецентральное интервальное оценивание и оценка статистических моделей.	2		4		17+1 (контроль)	24	основная: 1, 3 дополнительная: 2	<i>лекция-визуализация, проблемное обучение, обучение на основе опыта</i>
2	Байесовская классификация: Модель наивного байесовского классификатора. Оценка параметров для наивных байесовских моделей методом максимального правдоподобия. Построение классификатора по вероятностной модели. Апостериорное правило принятия решения (MAP). Применение наивного байесовского классификатора для задач фильтрации.	4		4	1	25+2 (контроль)	36	основная: 1, 2, 3 дополнительная: 1	<i>лекция-визуализация, проблемное обучение, обучение на основе опыта</i>
3	Анализ соответствий. Частотные таблицы сопряженности. Максимальная размерность. Координаты строк и столбцов в двумерном пространстве. Совместимость координат строк и столбцов. Шкалирование координат (возмож-	2	2	4	1	28+3 (контроль)	40	основная: 2 дополнительная: 1	<i>лекция-визуализация, проблемное обучение, обучение на</i>

	ности стандартизации). Метрика координатной системы. Оценка качества решения. Относительная инерция. Качество представления дополнительных точек. Многомерный анализ соответствий (МАС). Бинарная или индикаторная матрица сопряженности. Нечеткое кодирование. Матрица Берта.								<i>основе опыта</i>
4	Дискриминантный анализ. Выбор правил дискриминации. Параметрические и непараметрические методы дискриминации. Предпосылки параметрического дискриминантного анализа. Канонический дискриминантный анализ Линейный дискриминантный анализ Фишера. Статистика Уилкса и расстояние Махалобиса как критерии отбора информативных дискриминантных переменных.	4	2	4	1	30+3 (контроль)	54	основная: 3 дополнительная: 1, 2	<i>лекция-визуализация, проблемное обучение, обучение на основе опыта</i>

Занятия, проводимые в интерактивной форме, составляют 100% от общего количества аудиторных часов по дисциплине Прикладная эконометрика (продвинутый уровень).

Лабораторные работы

№ занятия	№ раздела	Тема	Кол-во часов
1	1	Анализ мощности выборки	4
2	2	Фильтрация на основе наивного байесовского классификатора	4
3	3	Проведение анализа сопряженности на основе частотных таблиц	4
4	4	Проведение дискриминантного анализа	4

Практические занятия

№ занятия	№ раздела	Тема	Кол-во часов
1	1	Представление таблиц соответствий в различных видах	2
2	2	Расчет расстояния Махаланобиса для отбора информативных дискриминантных переменных	2

4. Учебно-методическое обеспечение самостоятельной работы студентов

Тема 1 Анализ мощности выборки: .

Вопросы для самостоятельного изучения (подготовке к обсуждению):

1. Формирование критериев анализа мощности с учетом повторного отбора выборок.

Расчетные задания (задачи и пр.):

2. Провести нецентральное интервальное оценивание и оценку статистических моделей.

Тема 2 Байесовская классификация.

Изучить методы Data mining, основанные на байесовской классификации. Определить задачи для которых эти методы используются.

Тема 3 Анализ соответствий.

Вопросы для самостоятельного изучения (подготовке к обсуждению):

1. Нечеткое кодирование в анализе соответствий.
2. Шкалирование координат (возможности стандартизации) в анализе соответствий..

Расчетные задания (задачи и пр.) Методические указания даны в источнике основной литературы 1

Оценить качество представления дополнительных точек в анализе соответствий

Тема 4 Дискриминантный анализ:.

Вопросы для самостоятельного изучения (подготовке к обсуждению):

Линейный дискриминантный анализ Фишера

Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература

1. Тюрин, Ю. Н. Анализ данных на компьютере : [учебное пособие по направлениям "Математика", "Математика. Прикладная математика"] / Ю. Н. Тюрин, А. А. Макаров .— Изд. 4-е, перераб. — Москва : ИД "ФОРУМ", 2010 .— 367 с.
2. Халафян, А. А. STATISTICA 6. Статистический анализ данных : [учебник для студентов вузов, обучающихся по специальности "Статистика" и другим экономическим специальностям] / А. А. Халафян .— 2-е изд., [перераб. и доп.] .— Москва : БИНОМ, 2010 .— 522 с..
3. Сидняев, Н. И. Теория планирования эксперимента и анализ статистических данных : учебное пособие для магистров / Н. И. Сидняев .— 2-е изд., перераб. и доп. — Москва : Юрайт, 2015 .— 495 с.

Дополнительная литература

1. Кабаков, Р. R в действии. Анализ и визуализация данных на языке R / Р. И. Кабаков ; пер. с англ. П. А. Волковой .— Москва : ДМК Пресс, 2014 .— 587 с.
2. Куликов, Е. И. Прикладной статистический анализ : [учебное пособие для студентов высших учебных заведений, обучающихся по специальности 090106 - "Информационная безопасность телекоммуникационных систем", и аспирантов, обучающихся по специальности научных работников 05.13.19 - "Методы и системы защиты информации, информационная безопасность"] / Е. И. Куликов .— 2-е изд., перераб. и доп. — М. : Горячая линия-Телеком, 2008

Интернет-ресурсы (электронные учебно-методические издания, лицензионное программное обеспечение)

Neicon [Электронный ресурс]: архив научных журналов / Министерство образования и науки Российской Федерации; Национальный электронно-информационный консорциум (Neicon) - [Москва]: Нэйкон, 2015.

ScienceDirect. MATHEMATICS [Электронный ресурс]: тематическая полнотекстовая коллекция научных журналов / Издательство "Elsevier" - [Амстердам]: Elsevier, 2015

Образовательные технологии

При реализации дисциплины применяются классические образовательные технологии. При реализации дисциплины применяются интерактивные формы проведения лабораторных занятий в виде проблемного обучения. Проблемное обучение ориентировано на то что, магистрант всегда работает с реальными данными из области IT-сферы, что требует от него адаптации собственных знаний по дисциплине, возможно, в том числе за счет их самостоятельного расширения, для решения конкретной задачи анализа статистических данных.

Методические указания по освоению дисциплины

Раздел 1 Анализ мощности выборки

Лекций – 2 ч., лабораторное занятие – 4 часа, СРС – 17 ч., контроль (подготовка к зачету) – 1 час,

Магистрант должен иметь представление о методах формирования выборок, должен уметь определять их необходимый объем в зависимости от цели проводимого исследования. Магистрант должен иметь представление об отличиях и особенностях простой случайной, стратификационной и серийной выборки. Уметь понимать для решения каких задач, какая именно выборка данных

необходима. Магистрант должен иметь опыт проведения анализа мощности статистических критериев, использовать критерий значимости (RS/AS). Магистрант должен знать и уметь применять на практике углубленные методы доверительного интервального оценивания. Магистрант должен иметь использовать графические и аналитические процедуры анализа мощности для лучшей интерпретации результатов. Магистрант должен иметь использовать углубленные методы оценивания величины экспериментальных эффектов, проводить нецентральное интервальное оценивание и оценивать статистические модели. Для закрепления навыков применения анализа мощности магистранту предлагается выполнить задание № 1 из ФОС на лабораторной работе. Преподаватель на лабораторных занятиях объясняет лишь ход выполнения работы в ППП Statistica или R.

Раздел 2. Байесовская классификация

Лекций – 4 ч., лабораторное занятие – 4 часа, СРС – 25 ч., контроль (подготовка к зачету) – 2 часа, КСР -1 час.

В разделе изучаются модели наивного байесовского классификатора. Магистрант должен уметь оценивать параметры для наивных байесовских моделей методом максимального правдоподобия. Магистрант должен иметь опыт построения классификатора по вероятностной модели. Магистрант должен уметь рассчитывать апостериорное правило принятия решения (MAP). В результате изучения раздела у магистранта должен сформироваться опыт применения наивного байесовского классификатора для задач фильтрации.. Рекомендуется в качестве закрепления навыков построения модели наивного байесовского классификатора выполнить задание №2 из ФОС на лабораторной работе. Данные для выполнения задания магистрант выбирает самостоятельно, или преподаватель выдает ему данные. Преподаватель на лабораторных занятиях объясняет лишь ход выполнения работы в ППП Statistica или R.

Раздел 3. Анализ соответствий.

Лекций – 2 ч., лабораторное занятие – 4 часа, практические занятия – 2 часа, СРС – 28 ч., контроль (подготовка к зачету) – 3 часа, КСР -1 час.

Магистрант должен уметь строить частотные таблицы сопряженности. Магистрант должен уметь определять максимальную размерность для решения задач корреспонденций. Магистрант должен уметь правильно представлять координаты строк и столбцов в двумерном пространстве. Магистрант должен знать и уметь применять методы совместимости координат строк и столбцов. Магистрант должен уметь проводить шкалирование координат, с учетом возможности стандартизации. Магистрант должен иметь представление о метриках координатной системы. Магистрант должен уметь оценивать качество решения. Магистрант должен уметь определять общую и относительную инерцию. Магистрант должен уметь проводить анализ соответствий с введением дополнительных точек и оценивать качество представления дополнительных точек. Магистрант должен знать и уметь применять многомерный анализ соответствий (MAC). Магистрант должен знать принципы и уметь рассчитывать бинарную или индикаторную матрицу сопряженности. Магистрант должен знать и уметь проводить нечеткое кодирование в анализе соответствий, рассчитывать матрицу Берта. Рекомендуется в качестве закрепления навыков проведения анализа соответствий, выполнить задание №3 из ФОС на лабораторной работе. Для закрепления навыков практического счета таблиц соответствий в идее индикативной матрицы и матрицы Берта магистрантцу рекомендуется выполнить практическое задание №1 на практическом занятии. Данные для выполнения задания магистрант выбирает самостоятельно, или преподаватель выдает ему данные.

Раздел 4 Дискриминантный анализ.

Лекций – 4 ч., практические занятия – 2 часа, лабораторное занятие – 4 часа, СРС – 30 ч., контроль (подготовка к зачету) – 3 часа, КСР -1 час.

Магистрант должен иметь представление о параметрических и непараметрических методах дискриминации, предпосылки параметрического дискриминантного анализа. Уметь проводить выбор правил дискриминации. Магистрант должен освоить навыки применения канонического дискриминантного анализа и линейного дискриминантного анализа Фишера. Магистрант должен уметь применять статистику Уилкса и расстояние Махаланобиса в качестве критериев отбора информативных дискриминантных переменных. В качестве закрепления навыков проведения дискриминантного анализа, а также формирования умений его применений рекомендуется проведение лабораторной работы № 4 на данную тему, задание к которой отражено в ФОС. Алгоритм проведения дискриминантного анализа следующий: сначала понять какие независимые переменные следует анализировать, какой корень больше всего нагружают эти переменные, а затем посмотреть какая зависимая переменная нагружает этот же корень. Интерпретацию коэффициентов канонических функций можно смотреть и по абсолютному значению. Таким образом, рассматриваем наибольшие и наименьшие значения корней по модулю и формулируем правила, что чем больше абсолютные значения по этим переменным и меньше абсолютные значения по другим переменным, тем с большим успехом можно сказать, что наблюдение относится к определенному классу. Для закрепления навыков использования расстояния Махаланобиса в задаче определения принадлежности заданной точки одному из N классов в дискриминантном анализе рекомендуется выполнить задание для практического занятия №2. Преподаватель на лабораторных занятиях объясняет лишь ход выполнения работы в ППП Statistica или R.

Материально-техническое обеспечение дисциплины

лекционные аудитории с современными средствами демонстрации 6-415, 6-416, 6-213.

- кафедральные лаборатории, обеспечивающих реализацию ОПОП ВО: 6-218 Учебно-научная лаборатория «Технологии искусственного интеллекта в социально-экономических исследованиях, 6-417 Лаборатория информатики и программирования, 6-417а Учебно-научная лаборатория «Интеллектуальных технологий проектирования сложных систем».

Адаптация рабочей программы для лиц с ОВЗ

Адаптированная программа разрабатывается при наличии заявления со стороны обучающегося (родителей, законных представителей) и медицинских показаний (рекомендациями психолого-медико-педагогической комиссии). Для инвалидов адаптированная образовательная программа разрабатывается в соответствии с индивидуальной программой реабилитации.