

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Институт вычислительной математики и информационных технологий



УТВЕРЖДАЮ

Проректор по образовательной деятельности КФУ
проф. Таюрский Д.А.

"__" _____ 20__ г.

Программа дисциплины

Интеллектуальный анализ данных

Направление подготовки: 09.04.03 - Прикладная информатика

Профиль подготовки: Информационная безопасность экономических систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Содержание

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО
2. Место дисциплины (модуля) в структуре ОПОП ВО
3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся
4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий
 - 4.1. Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)
 - 4.2. Содержание дисциплины (модуля)
5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)
6. Фонд оценочных средств по дисциплине (модулю)
7. Перечень литературы, необходимой для освоения дисциплины (модуля)
8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)
9. Методические указания для обучающихся по освоению дисциплины (модуля)
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)
12. Средства адаптации преподавания дисциплины (модуля) к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья
13. Приложение №1. Фонд оценочных средств
14. Приложение №2. Перечень литературы, необходимой для освоения дисциплины (модуля)
15. Приложение №3. Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Программу дисциплины разработал(а)(и) ассистент, б.с. Нигматуллин Р.Р. (кафедра системного анализа и информационных технологий, отделение фундаментальной информатики и информационных технологий), RuslRNigmatullin@kpfu.ru

1. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения ОПОП ВО

Обучающийся, освоивший дисциплину (модуль), должен обладать следующими компетенциями:

Шифр компетенции	Расшифровка приобретаемой компетенции
ОК-1	способность к абстрактному мышлению, анализу, синтезу
ОК-3	способность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3	способность исследовать современные проблемы и методы прикладной информатики и научно-технического развития информационно-коммуникационных технологий
ОПК-4	способность исследовать закономерности становления и развития информационного общества в конкретной прикладной области
ОПК-5	способность на практике применять новые научные принципы и методы исследований
ПК-1	способность использовать и развивать методы научных исследований и инструментария в области проектирования и управления информационными системами в прикладных областях
ПК-14	способность принимать эффективные проектные решения в условиях неопределенности и риска
ПК-15	способность формировать стратегию информатизации прикладных процессов и создания прикладных ИС в соответствии со стратегией развития предприятий
ПК-21	способность использовать передовые методы оценки качества, надежности и информационной безопасности ИС в процессе эксплуатации прикладных ИС
ПК-8	способность анализировать данные и оценивать требуемые знания для решения нестандартных задач с использованием математических методов и методов компьютерного моделирования

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

о современных методах интеллектуального анализа данных.

Должен уметь:

понимать основные проблемы, возникающие при анализе данных, и пути их решения.

Должен владеть:

навыками анализа данных различной природы

Должен демонстрировать способность и готовность:

применять в своей дальнейшей профессиональной деятельности полученные знания и навыки

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.4 Дисциплины (модули)" основной профессиональной образовательной программы 09.04.03 "Прикладная информатика (Информационная безопасность экономических систем)" и относится к дисциплинам по выбору.

Осваивается на 2 курсе в 3 семестре.

3. Объем дисциплины (модуля) в зачетных единицах с указанием количества часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зачетных(ые) единиц(ы) на 180 часа(ов).

Контактная работа - 42 часа(ов), в том числе лекции - 14 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 28 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 84 часа(ов).

Контроль (зачёт / экзамен) - 54 часа(ов).

Форма промежуточного контроля дисциплины: экзамен в 3 семестре.

4. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

4.1 Структура и тематический план контактной и самостоятельной работы по дисциплине (модулю)

N	Разделы дисциплины / модуля	Семестр	Виды и часы контактной работы, их трудоемкость (в часах)			Самостоятельная работа
			Лекции	Практические занятия	Лабораторные работы	
1.	Тема 1. Введение в интеллектуальный анализ данных. Предварительная обработка данных.	3	2	0	4	12
2.	Тема 2. Линейные модели регрессии.	3	4	0	8	18
3.	Тема 3. Задача классификации с обучением. Логистическая регрессия.	3	4	0	8	18
4.	Тема 4. Нейронные сети для решения задач классификации с обучением.	3	2	0	4	18
5.	Тема 5. Кластерный анализ.	3	2	0	4	18
	Итого		14	0	28	84

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в интеллектуальный анализ данных. Предварительная обработка данных.

Области использования.

Информационный поиск (ранжирование веб-страниц)

Системы рекомендаций (реклама)

Анализ рыночной корзины

Обнаружение мошенничества (fraud)

Анализ лог-файлов

Идентификация сообществ в соц. сетях

Анализ тональности сообщений (сентимент анализ)

Связь с другими дисциплинами. Машинное обучение. Статистика. Теория информации

Компоненты систем Data Mining.

Предварительная обработка

Интеллектуальный анализ данных

Обработка выявленных закономерностей

Тема 2. Линейные модели регрессии.

Регрессия:

- Постановка задачи

- Модели регрессии

- Матрица плана (design matrix)

- Функция правдоподобия для задачи регрессии

- Целевая функция задачи регрессии

- Вывод аналитического решения задачи регрессии
- Понятие overfitting и underfitting
- Регуляризация
- Целевая функция для задачи регрессии с регуляризацией
- Вывод аналитического решения задачи регрессии с регуляризацией
- Построение модели на основе разделения данных на обучающую, валидационную и тестовую выборки

Тема 3. Задача классификации с обучением. Логистическая регрессия.

Классификация:

- Постановка задачи
- Метрики для бинарной задачи классификации: TP, TN, FP, FN, Ошибка первого рода, Ошибка второго рода, Accuracy, Precision, Recall, F1 Score
- Метрики для задачи мультиклассовой классификации: Матрица неточностей (confusion matrix), Accuracy
- Логистическая регрессия
- Понятие классификатора на основе логистической регрессии для решения задачи бинарной классификации
- Функция правдоподобия для логистической регрессии
- Целевая функция логистической регрессии
- Градиент целевой функции логистической регрессии
- Метод градиентного спуска для решения задачи логистической регрессии
- Регуляризация логистической регрессии
- Softmax регрессия
- one-hot encoding
- функция softmax и её свойства
- Целевая функция softmax регрессии

Тема 4. Нейронные сети для решения задач классификации с обучением.

Многослойный перцептрон

- Архитектура
- Понятие нейрона, слоя. Входной и выходной слой
- Матрица весов
- Функция активации (HardLim, Linear, Sigmoid, Tanh, ReLU)
- Распространение сигнала через нейрон.
- Выходной слой для задачи регрессии и классификации
- Целевые функции
- задача регрессии
- задача классификации
- Обучение
- Стохастический градиентный спуск (SGD)
- Разделение целевой функции
- Понятие минибатча
- Итерации и эпохи
- Критерии выхода из процесса обучения
- Алгоритм прямого распространения ошибки и обратного распространения сигнала (Feed forward back propagation)
- Инициализация параметров MLP для обучения
- Нормализация данных

Тема 5. Кластерный анализ.

Кластеризация

- Постановка задачи
- Примеры расстояний
- Классификация алгоритмов кластеризации
- Алгоритм k-means
- Иерархическая кластеризация
- Понятие расстояния между кластерами возможные определения
- Агломеративная кластеризация
- Кластеризация на основе графового представления данных

- использование минимального остоного дерева
- использование порога
- Алгоритм c-means

5. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине (модулю)

Самостоятельная работа обучающихся выполняется по заданию и при методическом руководстве преподавателя, но без его непосредственного участия. Самостоятельная работа подразделяется на самостоятельную работу на аудиторных занятиях и на внеаудиторную самостоятельную работу. Самостоятельная работа обучающихся включает как полностью самостоятельное освоение отдельных тем (разделов) дисциплины, так и проработку тем (разделов), осваиваемых во время аудиторной работы. Во время самостоятельной работы обучающиеся читают и конспектируют учебную, научную и справочную литературу, выполняют задания, направленные на закрепление знаний и отработку умений и навыков, готовятся к текущему и промежуточному контролю по дисциплине.

Организация самостоятельной работы обучающихся регламентируется нормативными документами, учебно-методической литературой и электронными образовательными ресурсами, включая:

Порядок организации и осуществления образовательной деятельности по образовательным программам высшего образования - программам бакалавриата, программам специалитета, программам магистратуры (утвержден приказом Министерства образования и науки Российской Федерации от 5 апреля 2017 года №301)

Письмо Министерства образования Российской Федерации №14-55-996ин/15 от 27 ноября 2002 г. "Об активизации самостоятельной работы студентов высших учебных заведений"

Устав федерального государственного автономного образовательного учреждения "Казанский (Приволжский) федеральный университет"

Правила внутреннего распорядка федерального государственного автономного образовательного учреждения высшего профессионального образования "Казанский (Приволжский) федеральный университет"

Локальные нормативные акты Казанского (Приволжского) федерального университета

6. Фонд оценочных средств по дисциплине (модулю)

Фонд оценочных средств по дисциплине (модулю) включает оценочные материалы, направленные на проверку освоения компетенций, в том числе знаний, умений и навыков. Фонд оценочных средств включает оценочные средства текущего контроля и оценочные средства промежуточной аттестации.

В фонде оценочных средств содержится следующая информация:

- соответствие компетенций планируемым результатам обучения по дисциплине (модулю);
- критерии оценивания сформированности компетенций;
- механизм формирования оценки по дисциплине (модулю);
- описание порядка применения и процедуры оценивания для каждого оценочного средства;
- критерии оценивания для каждого оценочного средства;
- содержание оценочных средств, включая требования, предъявляемые к действиям обучающихся, демонстрируемым результатам, задания различных типов.

Фонд оценочных средств по дисциплине находится в Приложении 1 к программе дисциплины (модулю).

7. Перечень литературы, необходимой для освоения дисциплины (модуля)

Освоение дисциплины (модуля) предполагает изучение основной и дополнительной учебной литературы.

Литература может быть доступна обучающимся в одном из двух вариантов (либо в обоих из них):

- в электронном виде - через электронные библиотечные системы на основании заключенных КФУ договоров с правообладателями;

- в печатном виде - в Научной библиотеке им. Н.И. Лобачевского. Обучающиеся получают учебную литературу на абонементе по читательским билетам в соответствии с правилами пользования Научной библиотекой.

Электронные издания доступны дистанционно из любой точки при введении обучающимся своего логина и пароля от личного кабинета в системе "Электронный университет". При использовании печатных изданий библиотечный фонд должен быть укомплектован ими из расчета не менее 0,5 экземпляра (для обучающихся по ФГОС 3++ - не менее 0,25 экземпляра) каждого из изданий основной литературы и не менее 0,25 экземпляра дополнительной литературы на каждого обучающегося из числа лиц, одновременно осваивающих данную дисциплину.

Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля), находится в Приложении 2 к рабочей программе дисциплины. Он подлежит обновлению при изменении условий договоров КФУ с правообладателями электронных изданий и при изменении комплектования фондов Научной библиотеки КФУ.

8. Перечень ресурсов информационно-телекоммуникационной сети "Интернет", необходимых для освоения дисциплины (модуля)

Интернет-портал по алгоритмике и программированию - <http://algolist.manual.ru/>

Интернет-портал по технологии анализа данных - <http://www.olap.ru/>

Интернет-портал ресурсов по ИТ - <http://www.intuit.ru>

Интернет-портал ресурсов по математическим наукам - <http://www.math.ru/>

Электронная библиотека по техническим наукам - <http://techlibrary.ru>

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Вид работ	Методические рекомендации
лекции	<p>Во время занятий студенты должны сосредоточить внимание на рассматриваемом материале. Основные положения, важные определения и теоретические положения необходимо записывать. Конспектирование предлагаемого преподавателем материала вырабатывает у студентов навыки самостоятельного отбора и анализа необходимой для них информации, умение более сжато и четко записывать услышанное. Необходимо добиваться полного понимания излагаемого на занятии материала. В случае возникновения неясностей или недопонимания, необходимо задавать вопросы преподавателю. Особое внимание следует уделить выполнению заданий и упражнений, предлагаемых преподавателем, которые служат закреплению усвоения рассматриваемой темы.</p> <p>Выполнение примеров помогает добиться правильного понимания материала. Конспекты могут служить необходимым вспомогательным материалом в процессе подготовки к экзамену.</p>
лабораторные работы	<p>Данная дисциплина включает лабораторные занятия, на которых теоретический материал, полученный на лекциях закрепляется в виде практических упражнений. Это позволяет студентам глубже усвоить материал и лучше понять детали лекционного материала. Лабораторные работы предполагают получение практических навыков по решению задач анализа данных. При этом в процессе работы студенту необходимо активно использовать записи лекций и прикладные материалы.</p>
самостоятельная работа	<p>Изучение данной дисциплины предусматривает систематическую самостоятельную работу студентов над дополнительными материалами; развитие навыков самоконтроля, способствующих интенсификации учебного процесса. Изучение лекционного материала по конспекту лекций должно сопровождаться изучением рекомендуемой литературы, основной и дополнительной. Основной целью организации самостоятельной работы студентов является систематизация и активизация знаний, полученных ими на лекциях и в процессе подготовки к практическим занятиям. Студентам следует стремиться к активизации знаний на занятиях. Практические занятия посвящены получению навыков решения прикладных задач анализа данных. Здесь магистранту следует активно работать с конспектом лекций для решения поставленных перед ними задач.</p> <p>Самостоятельная работа по изучению курса 'Интеллектуальный анализ данных' предполагает внеаудиторную работу, которая включает:</p> <ol style="list-style-type: none"> 1. Просмотр записей курса, повторение и закрепление материала. 2. Выполнение упражнений, предложенных для самостоятельного решения. 3. Изучение вопросов, оставленных на самостоятельное изучение. 4. Подготовку к экзамену.

Вид работ	Методические рекомендации
экзамен	Процесс подготовки к экзамену служит систематизации знаний, полученных в течение семестра при изучении данного курса. При подготовке к экзамену магистрант должен воспользоваться конспектами лекций, сделанных им в течении семестра. В случае недостаточно хорошего изложения материала в лекциях, в случае вопросов и непонимания отдельных моментов, а также при рассмотрении тем, вынесенных на самостоятельную работу студента, необходимо воспользоваться литературой из списка основной и дополнительной литературы. Особо важным этапом является резюме прочитанного теоретического источника, так как это является важным условием подготовки к экзамену. Также необходимо еще раз вернуться к тем примерам и упражнениям, которые рассматривались на занятиях, а также предлагались для самостоятельного выполнения. В результате подготовки к экзамену студент должен иметь не обрывочные знания по отдельным темам курса, а обладать полной картиной, соответствующей изученной дисциплине.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем (при необходимости)

Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения и информационных справочных систем, представлен в Приложении 3 к рабочей программе дисциплины (модуля).

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю)

Материально-техническое обеспечение образовательного процесса по дисциплине (модулю) включает в себя следующие компоненты:

Помещения для самостоятельной работы обучающихся, укомплектованные специализированной мебелью (столы и стулья) и оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду КФУ.

Учебные аудитории для контактной работы с преподавателем, укомплектованные специализированной мебелью (столы и стулья).

Компьютер и принтер для распечатки раздаточных материалов.

Мультимедийная аудитория.

Компьютерный класс.

12. Средства адаптации преподавания дисциплины к потребностям обучающихся инвалидов и лиц с ограниченными возможностями здоровья

При необходимости в образовательном процессе применяются следующие методы и технологии, облегчающие восприятие информации обучающимися инвалидами и лицами с ограниченными возможностями здоровья:

- создание текстовой версии любого нетекстового контента для его возможного преобразования в альтернативные формы, удобные для различных пользователей;
- создание контента, который можно представить в различных видах без потери данных или структуры, предусмотреть возможность масштабирования текста и изображений без потери качества, предусмотреть доступность управления контентом с клавиатуры;
- создание возможностей для обучающихся воспринимать одну и ту же информацию из разных источников - например, так, чтобы лица с нарушениями слуха получали информацию визуально, с нарушениями зрения - аудиально;
- применение программных средств, обеспечивающих возможность освоения навыков и умений, формируемых дисциплиной, за счёт альтернативных способов, в том числе виртуальных лабораторий и симуляционных технологий;
- применение дистанционных образовательных технологий для передачи информации, организации различных форм интерактивной контактной работы обучающегося с преподавателем, в том числе вебинаров, которые могут быть использованы для проведения виртуальных лекций с возможностью взаимодействия всех участников дистанционного обучения, проведения семинаров, выступления с докладами и защиты выполненных работ, проведения тренингов, организации коллективной работы;
- применение дистанционных образовательных технологий для организации форм текущего и промежуточного контроля;
- увеличение продолжительности сдачи обучающимся инвалидом или лицом с ограниченными возможностями здоровья форм промежуточной аттестации по отношению к установленной продолжительности их сдачи;
- продолжительности сдачи зачёта или экзамена, проводимого в письменной форме, - не более чем на 90 минут;

- продолжительности подготовки обучающегося к ответу на зачёте или экзамене, проводимом в устной форме, - не более чем на 20 минут;
- продолжительности выступления обучающегося при защите курсовой работы - не более чем на 15 минут.

Программа составлена в соответствии с требованиями ФГОС ВО и учебным планом по направлению 09.04.03 "Прикладная информатика" и магистерской программе "Информационная безопасность экономических систем".

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.04.03 - Прикладная информатика

Профиль подготовки: Информационная безопасность экономических систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Основная литература:

1. Шапиро, Л. Компьютерное зрение [Электронный ресурс] / Л. Шапиро, Д. Стокман. - Электрон. дан. - Москва : Издательство 'Лаборатория знаний', 2015. - 763 с. - Режим доступа: <https://e.lanbook.com/book/84096>.
2. Жданов, А.А. Автономный искусственный интеллект [Электронный ресурс] : учеб. пособие - Электрон. дан. - Москва : Издательство 'Лаборатория знаний', 2015. - 362 с. - Режим доступа: <https://e.lanbook.com/book/70761>.
3. Вьюгин, В.В. Математические основы машинного обучения и прогнозирования. [Электронный ресурс] - Электрон. дан. - М. : МЦНМО (Московский центр непрерывного математического образования), 2014. - 304 с. - Режим доступа: http://e.lanbook.com/books/element.php?pl1_id=56397

Дополнительная литература:

1. Информационные аналитические системы [Электронный ресурс] : учебник / Т. В. Алексеева, Ю. В. Амириди, В. В. Дик и др.; под ред. В. В. Дика. - М.: МФПУ Синергия, 2013. - 384 с. - Режим доступа: <http://znanium.com/bookread2.php?book=451186>
2. Осипов Г. В. Методы искусственного интеллекта/Осипов Г.В. - М.: Физматлит, 2011. - 296 с. - Режим доступа: <http://znanium.com/bookread2.php?book=544787>
3. Введение в специальность программиста: Учебник / В.А. Гвоздева. - 2-е изд., испр. и доп. - М.: ИД ФОРУМ: НИЦ Инфра-М, 2013. - 208 с. - Режим доступа: <http://znanium.com/bookread2.php?book=398911>

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.4 Интеллектуальный анализ данных

Перечень информационных технологий, используемых для освоения дисциплины (модуля), включая перечень программного обеспечения и информационных справочных систем

Направление подготовки: 09.04.03 - Прикладная информатика

Профиль подготовки: Информационная безопасность экономических систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2017

Освоение дисциплины (модуля) предполагает использование следующего программного обеспечения и информационно-справочных систем:

Операционная система Microsoft Windows 7 Профессиональная или Windows XP (Volume License)

Пакет офисного программного обеспечения Microsoft Office 365 или Microsoft Office Professional plus 2010

Браузер Mozilla Firefox

Браузер Google Chrome

Adobe Reader XI или Adobe Acrobat Reader DC

Kaspersky Endpoint Security для Windows

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе "ZNANIUM.COM", доступ к которой предоставлен обучающимся. ЭБС "ZNANIUM.COM" содержит произведения крупнейших российских учёных, руководителей государственных органов, преподавателей ведущих вузов страны, высококвалифицированных специалистов в различных сферах бизнеса. Фонд библиотеки сформирован с учетом всех изменений образовательных стандартов и включает учебники, учебные пособия, учебно-методические комплексы, монографии, авторефераты, диссертации, энциклопедии, словари и справочники, законодательно-нормативные документы, специальные периодические издания и издания, выпускаемые издательствами вузов. В настоящее время ЭБС ZNANIUM.COM соответствует всем требованиям федеральных государственных образовательных стандартов высшего образования (ФГОС ВО) нового поколения.

Учебно-методическая литература для данной дисциплины имеется в наличии в электронно-библиотечной системе Издательства "Лань", доступ к которой предоставлен обучающимся. ЭБС Издательства "Лань" включает в себя электронные версии книг издательства "Лань" и других ведущих издательств учебной литературы, а также электронные версии периодических изданий по естественным, техническим и гуманитарным наукам. ЭБС Издательства "Лань" обеспечивает доступ к научной, учебной литературе и научным периодическим изданиям по максимальному количеству профильных направлений с соблюдением всех авторских и смежных прав.