

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение высшего образования
"Казанский (Приволжский) федеральный университет"
Высшая школа информационных технологий и интеллектуальных систем



УТВЕРЖДАЮ
Проректор по образовательной деятельности КФУ
Проф. Д. А. Таюрский

» _____ 20__ г.

подписано электронно-цифровой подписью

Программа дисциплины

Интеллектуальный анализ данных

Направление подготовки: 09.04.04 - Программная инженерия

Профиль подготовки: Разработка программно-информационных систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

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) доцент, к.н. (доцент) Григорян К.А. (кафедра Цифровые технологии в экономике (базовая), Высшая школа информационных технологий и интеллектуальных систем), KAGrigoryan@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 |
|------------------|---|
| ПК-5 | Способность применять перспективные методы исследования и решения профессиональных задач на основе знания мировых тенденций развития вычислительной техники и телекоммуникационных технологий |

Обучающийся, освоивший дисциплину (модуль):

Должен демонстрировать способность и готовность:

Строить математические модели и применять методы интеллектуального анализа данных в бизнесе для принятия стратегических и тактических решений,

применять методы машинного обучения и владеть актуальными для индустрии аналитическими программными пакетами.

2. Место дисциплины (модуля) в структуре ОПОП ВО

Данная дисциплина (модуль) включена в раздел "Б1.В.ДВ.04.01 Дисциплины (модули)" основной профессиональной образовательной программы 09.04.04 "Программная инженерия (Разработка программно-информационных 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 4 зачетных(ые) единиц(ы) на 144 часа(ов).

Контактная работа - 36 часа(ов), в том числе лекции - 0 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 36 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 108 часа(ов).

Контроль (зачёт / экзамен) - 0 часа(ов).

Форма промежуточного контроля дисциплины: зачет 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. Интеллектуальный анализ данных и машинное обучение в условиях цифровой экономики | 3 | 0 | 0 | 4 | 10 |
| 2. | Тема 2. Методы и технологии интеллектуального анализа данных в бизнесе | 3 | 0 | 0 | 4 | 10 |
| 3. | Тема 3. Этапы процесса "data mining" и основы прогнозного моделирования | 3 | 0 | 0 | 4 | 18 |
| 4. | Тема 4. Задачи и алгоритмы построения и оптимизации "деревьев решений" | 3 | 0 | 0 | 8 | 20 |
| 5. | Тема 5. Сфера применения и особенность предсказаний регрессионных моделей | 3 | 0 | 0 | 8 | 20 |
| 6. | Тема 6. Введение в выявление закономерностей и кластерный анализ | 3 | 0 | 0 | 4 | 15 |
| 7. | Тема 7. Оценка, сравнение и применение моделей | 3 | 0 | 0 | 4 | 15 |
| | Итого | | 0 | 0 | 36 | 108 |

4.2 Содержание дисциплины (модуля)

Тема 1. Интеллектуальный анализ данных и машинное обучение в условиях цифровой экономики

Основные понятия data mining и машинного обучения. Виды данных. Принятие решений и конкуренция на основе аналитики в цифровой экономике. Аналитическая экономика. Аналитика и результаты бизнеса. Анализ данных в маркетинге. Прогноз привлечения и оттока клиентов. Кластеризация клиентов и анализ профилей кластера. Задачи выявления ассоциаций и последовательностей. Применение прогнозного моделирования в банковском скоринге и в задачах выявления мошенников. Текстовая аналитика и сферы ее применения.

Тема 2. Методы и технологии интеллектуального анализа данных в бизнесе

Применение машинного обучения в интеллектуальном анализе данных. Контролируемое обучение (с учителем). Неконтролируемое обучение (без учителя). Методы прогнозного моделирования. Методы поиска закономерностей (ассоциации и кластеризация). Прогнозирование новых значений. Выбор входных переменных для анализа и проблема размерности. Оптимизация сложности модели. Задачи, возможности и инструменты интеллектуального анализа данных в пакете SAS Enterprise Miner.

Тема 3. Этапы процесса "data mining" и основы прогнозного моделирования

Задачи прогнозных моделей. Характеристика этапов процесса "data mining". Назначение цели исследования и создание проектного задания. Сбор данных и проверка их качества. Очистка, интеграция и преобразование данных. Исследовательский анализ данных. Построение моделей. Диагностика и сравнение моделей. Применение модели и представление результатом.

Тема 4. Задачи и алгоритмы построения и оптимизации "деревьев решений"

Характеристика прогнозных моделей на основе "дерева решений". Создание обучающих и проверочных данных. Интерактивное построение дерева решений. Информационная энтропия. Создание и редактирование правил разбиения. Создание максимального дерева. Оптимизация сложности деревьев решений. Оценка качества дерева решений. "Обрубка" дерева решений и выбор оптимального дерева. Изучение и применение настроек автоматического создания дерева решений в пакете SAS Enterprise Miner.

Тема 5. Сфера применения и особенность предсказаний регрессионных моделей

Характеристика и особенность регрессионных моделей. Линейная и логистическая регрессия. Логит-функция. Проблема пропущенных значений и стратегии их замены. Обработка пропущенных значений в SAS Enterprise Miner. Запуск и настройка узла Regression. Выбор входных переменных регрессионной модели. Характеристика методов forward selection, backward selection и stepwise. Оптимизация сложности регрессии. Выбор лучшей модели в последовательности. Интерпретация регрессионных моделей. Преобразование входных переменных регрессионной модели. Работа с категориальными входными переменными. Полиномиальная регрессия.

Тема 6. Введение в выявление закономерностей и кластерный анализ

Основные понятия, возможности и проблемы выявления закономерностей в данных. Применение результатов кластеризации (сегментации) в решении задач бизнес-структур. Взаимосвязь кластеризации и прогнозного моделирования. Алгоритмы кластерного анализа. Алгоритм k-means. Этапы проведения кластерного анализа данных. Преобразование и стандартизация входных данных. Фильтрация данных и удаление экстремальных значений. Создание кластеров с помощью инструмента Cluster в пакете SAS Enterprise Miner. Выбор целевого числа сегментов кластера. Оценка параметров кластеризации и характеристика выделенных кластеров. Профилирование и описание сегментов.

Тема 7. Оценка, сравнение и применение моделей

Группировка статистик модели в зависимости от типа прогноза. Сравнение моделей с помощью сводных статистик. Сравнение моделей с помощью ROC- кривых. Сравнение моделей с помощью графиков рейтингов. Инструмент поправки на раздельную выборку в пакете SAS Enterprise Miner. Оценка и оптимизация прибыли для моделей. Встроенный скоринг наборов данных в пакете SAS Enterprise Miner. Экспорт и сохранение результатов скоринга. Применение модели вне среды SAS.

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)

Госкомстат РФ - <http://www.gks.ru>

Основные понятия интеллектуального анализа данных -

<https://docs.microsoft.com/ru-ru/sql/analysis-services/data-mining/data-mining-concepts?view=sql-server-2017#PreparingData>

Сайт содержит материалы по методам интеллектуально анализа данных - <https://statanaliz.info/>

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

| Вид работ | Методические рекомендации |
|------------------------|---|
| лабораторные работы | Основными видами контактной работы студентов являются лабораторные работы. В ходе лабораторных работ преподаватель излагает и разъясняет основные, наиболее сложные понятия темы, а также связанные с ней теоретические и практические проблемы, даёт рекомендации и указания на выполнение заданий и самостоятельную работу. |
| самостоятельная работа | Одним из основных видов деятельности обучающегося является самостоятельная работа, которая включает в себя изучение лекционного материала, учебников и учебных пособий, первоисточников, подготовку сообщений, выступления на групповых занятиях, выполнение индивидуальных домашних заданий. Методика самостоятельной работы предварительно разъясняется преподавателем и в последующем может уточняться с учетом индивидуальных особенностей обучающихся. Время и место самостоятельной работы выбираются обучающимися по своему усмотрению с учетом рекомендаций преподавателя. Самостоятельную работу над дисциплиной следует начинать с изучения программы, которая содержит основные требования к знаниям, умениям и навыкам обучаемых. Следует обязательно вспомнить рекомендации преподавателя, данные в ходе установочных занятий. Затем следует приступить к изучению отдельных разделов и тем в порядке, предусмотренном программой. |
| зачет | Зачёт нацелен на комплексную проверку освоения дисциплины. Обучающийся получает вопрос (вопросы) либо задание (задания) и время на подготовку. Зачёт проводится в устной, письменной или компьютерной форме. Оценивается владение материалом, его системное освоение, способность применять нужные знания, навыки и умения при анализе проблемных ситуаций и решении практических заданий. |

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 09.04.04 "Программная инженерия" и магистерской программе "Разработка программно-информационных систем".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.04.01 Интеллектуальный анализ данных

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 09.04.04 - Программная инженерия

Профиль подготовки: Разработка программно-информационных систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

Основная литература:

Много цифр. Анализ больших данных при помощи Excel / Форман Д.; Пер. с англ. Соколовой А. - М.: Альпина Пабли., 2016. - 461 с. ISBN 978-5-9614-5032-3 - Режим доступа: <http://znanium.com/catalog/product/551044>

Компьютерные технологии анализа данных в эконометрике: Монография / Д.М. Дайитбегов. - 3-е изд., испр. и доп. - М.: Вузовский учебник: НИЦ Инфра-М, 2013. - XIV, 587 с. (Научная книга). ISBN 978-5-9558-0275-6- Режим доступа: <http://znanium.com/catalog/product/365692>

Статистический анализ данных, моделирование и исследование вероятностных закономерностей. Компьютерный подход / Б.Ю. Лемешко, С.Б. Лемешко, С.Н. Постовалов и др. - М.: НИЦ ИНФРА-М, 2015. - 890 с. ISBN 978-5-16-103267-1 (online) - Режим доступа: <http://znanium.com/bookread2.php?book=515227>

Дополнительная литература:

Интеллектуальный анализ данных и систем управления бизнес-правилами в телекоммуникациях: Монография / Р.Р. Вейнберг. - М.: НИЦ ИНФРА-М, 2016. - 173 с. - (Научная мысль) ISBN 978-5-16-011350-0- Режим доступа: <http://znanium.com/bookread2.php?book=520998>

Статистический анализ данных в MS Excel : учеб. пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. ? М. : ИНФРА-М, 2019. ? 320 с. ? (Высшее образование). - Режим доступа: <http://znanium.com/catalog/product/987337>

Статистические методы анализа данных : учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.] ; под общ. ред. д-ра экон. наук, проф. Л.И. Ниворожкиной. ? М. : РИОР : ИНФРА-М, 2016. ? 333 с. ? (Высшее образование). - Режим доступа: <http://znanium.com/catalog/product/556760>

Приложение 3
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.04.01 Интеллектуальный анализ данных

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

Направление подготовки: 09.04.04 - Программная инженерия

Профиль подготовки: Разработка программно-информационных систем

Квалификация выпускника: магистр

Форма обучения: очное

Язык обучения: русский

Год начала обучения по образовательной программе: 2018

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.