

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

Технология Data Mining

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Математические основы и программное обеспечение информационной безопасности и защиты информации

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, к.н. Разинков Е.В. (кафедра системного анализа и информационных технологий, отделение фундаментальной информатики и информационных технологий), Evgenij.Razinkov@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	готовность к саморазвитию, самореализации, использованию творческого потенциала
ОПК-3	способность использовать и применять углубленные теоретические и практические знания в области фундаментальной информатики и информационных технологий
ОПК-4	способность самостоятельно приобретать и использовать в практической деятельности новые знания и умения, в том числе, в новых областях знаний, непосредственно не связанных со сферой деятельности, расширять и углублять свое научное мировоззрение
ОПК-5	способность использовать углубленные знания правовых и этических норм при оценке последствий своей профессиональной деятельности, при разработке и осуществлении социально значимых проектов
ПК-1	способность проводить научные исследования и получать новые научные и прикладные результаты самостоятельно и в составе научного коллектива
ПК-13	способность разрабатывать аналитические обзоры состояния области прикладной математики и информационных технологий
ПК-14	способность выполнять работу экспертов в ведомственных, отраслевых или государственных экспертных группах по экспертизе проектов, тематика которых соответствует направленности (профилю) программы магистратуры
ПК-16	способность участвовать в деятельности профессиональных сетевых сообществ по конкретным направлениям
ПК-2	способность использовать углубленные теоретические и практические знания в области информационных технологий и прикладной математики, фундаментальных концепций и системных методологий, международных и профессиональных стандартов в области информационных технологий
ПК-3	способность разрабатывать концептуальные и теоретические модели решаемых научных проблем и задач проектной и производственно-технологической деятельности
ПК-4	способность разрабатывать архитектурные и функциональные спецификации создаваемых систем и средств информационных технологий, а также разрабатывать абстрактные методы их тестирования
ПК-5	способность управлять проектами, планировать научно-исследовательскую деятельность, анализировать риски, управлять командой проекта
ПК-6	способность к углубленному анализу проблем, постановке и обоснованию задач научной и проектно-технологической деятельности

Обучающийся, освоивший дисциплину (модуль):

Должен знать:

о современных методах Data Mining.

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 02.04.02 "Фундаментальная информатика и информационные технологии (Математические основы и программное обеспечение информационной безопасности и защиты информации)" 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 - 28 часа(ов), в том числе лекции - 14 часа(ов), практические занятия - 0 часа(ов), лабораторные работы - 14 часа(ов), контроль самостоятельной работы - 0 часа(ов).

Самостоятельная работа - 62 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 Data Mining	3	4	0	4	9
2.	Тема 2. Методы Frequent Pattern Mining	3	2	0	2	9
3.	Тема 3. Методы анализа текста	3	2	0	2	11
4.	Тема 4. Вычисление векторных представлений слов	3	2	0	2	11
5.	Тема 5. Методы обнаружения аномалий	3	2	0	2	11
6.	Тема 6. Кластерный анализ	3	2	0	2	11
	Итого		14	0	14	62

4.2 Содержание дисциплины (модуля)

Тема 1. Введение в Data Mining

История Data Mining. Приложения, решаемые задачи: кластерный анализ, обнаружение аномалий, поиск часто встречающихся наборов элементов. Связь между машинным обучением и Data Mining. Классификация алгоритмов Data Mining. Актуальность методов. Основные сложности в обработке данных. Обзор источников информации по Data Mining,

Тема 2. Методы Frequent Pattern Mining

Постановка задачи Frequent Pattern Mining. Понятие часто встречающегося набора элементов. Метод Apriori. Недостатки метода Apriori: необходимость построения множества кандидатов. Метод FP-Growth: построение FP-дерева, анализ FP-дерева. Разбор применения метода Apriori и FP-Growth на примерах. Понятие Closed Pattern и Maximal Pattern.

Тема 3. Методы анализа текста

Приложения анализа текста. Представление текста в виде вектора. Модель Bag-of-Words, Построение словаря. Вычисление функции сходства между двумя документами. Методы Term Frequency и Inverted Document Frequency повышения точности вычисления функции сходства между текстовыми документами. Преимущества и недостатки модели Bag-of-Words.

Тема 4. Вычисление векторных представлений слов

Задача вычисления векторных представлений слов. Метод Word2Vec. Целевая функция метода Word2Vec. Архитектура нейронной сети в методе Word2Vec. Повышение эффективности обучения нейронной сети с помощью метода Negative Sampling. Свойства получаемых векторных представлений слов. Применение векторов Word2Vec.

Тема 5. Методы обнаружения аномалий

Задача обнаружения аномалий. Типы аномалий: точечные, контекстные, групповые аномалии. Приложения обнаружения аномалий. Сложность задачи обнаружения аномалий. Метод k-го ближайшего соседа. Эффективная реализация метода k-го ближайшего соседа. Недостатки метода k-го ближайшего соседа. Метод Local Outlier Factor.

Тема 6. Кластерный анализ

Задача кластеризации. Приложения кластерного анализа. Обзор методов кластеризации. Метод k-means. Агломеративная кластеризация. Метод кластеризации ROCK. Кластеризация с помощью деревьев решений. Преимущества и недостатки различных методов кластеризации. Применимость методов кластеризации в зависимости от особенностей данных.

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	Студенту рекомендуется внимательно слушать лектора, следить за тем, что написано на доске или представлено на слайдах презентации, анализировать получаемую им информацию. В случае, если материал лекции непонятен, следует задать вопрос в отведенное для вопросов время. Студенту также рекомендуется конспектировать материал лекции в тетради, что улучшает запоминание.
лабораторные работы	При выполнении лабораторных работ студенту рекомендуется внимательно анализировать поставленную задачу, уделяя особое внимание критериям оценки точности решения задачи. Программный код должен быть объектно-ориентированным, чистым, с поясняющими комментариями. Особое внимание следует уделить методологическим аспектам решения задачи, на корректное разделение выборки на обучающую, валидационную и тестовую. Результаты работы программы должны быть оформлены в виде таблиц и графиков.
самостоятельная работа	При ведении самостоятельной работы студенту рекомендуется внимательно подходить к изучению научных статей, обращать внимание на значимость полученного результата, на требования к обучающей выборке, на скорость работы предлагаемых алгоритмов, на результаты их сравнения с существующими. В случае, если изучаемый материал понятен не до конца, рекомендуется обращение к дополнительной литературе.
экзамен	Студенту рекомендуется внимательно анализировать вопросы в экзаменационном билете. Ответ на экзаменационный билет должен быть подробным и четким, все релевантные формулы должны быть приведены и пояснены. При ответе на вопрос студент должен проявить не столь умение запомнить материал, сколь глубокое его понимание. Рекомендуется избегать приведения в ответе материала, не относящегося к билету.

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 02.04.02 "Фундаментальная информатика и информационные технологии" и магистерской программе "Математические основы и программное обеспечение информационной безопасности и защиты информации".

Приложение 2
к рабочей программе дисциплины (модуля)
Б1.В.ДВ.4 Технология Data Mining

Перечень литературы, необходимой для освоения дисциплины (модуля)

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Математические основы и программное обеспечение информационной безопасности и защиты информации

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 Технология Data Mining

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

Направление подготовки: 02.04.02 - Фундаментальная информатика и информационные технологии

Профиль подготовки: Математические основы и программное обеспечение информационной безопасности и защиты информации

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.