

МИНОБРНАУКИ РОССИИ
Ярославский государственный университет им. П.Г. Демидова

Кафедра компьютерных сетей

УТВЕРЖДАЮ

Декан факультета ИВТ

 Д.Ю. Чалый

« 23 » мая 2023 г.

Рабочая программа дисциплины
«Технологии больших данных и Data Mining»

Направление подготовки
01.04.02 Прикладная математика и информатика

Направленность (профиль)
«Математические основы искусственного интеллекта»

Квалификация выпускника
Магистр

Форма обучения
очная

Программа рассмотрена
на заседании кафедры
от «17» апреля 2023 г.,
протокол № 8

Программа одобрена НМК
факультета ИВТ
протокол № 6 от
«28» апреля 2023 г.

Ярославль

1. Цели освоения дисциплины

Дисциплина «Технологии больших данных и Data Mining» предназначена для ознакомления магистрантов с основами одного из современных направлений в области обработки информации в хранилищах данных: технологии интеллектуального (Data Mining), визуального (Visual Mining), и текстового (Text Mining) анализа. Основная задача курса состоит в том, чтобы выпускники магистратуры приобрели необходимый минимум знаний в практически важной и активно развивающейся области прикладной информатики и получили бы навыки в решении возникающих в этой области задач

2. Место дисциплины в структуре ОП магистратуры

Дисциплина «Технологии больших данных и Data Mining» относится к вариативной части (дисциплина по выбору) ОП магистратуры. Она входит в список дисциплин по выбору вариативной части математического и естественнонаучного цикла

Дисциплина относится к числу прикладных направлений информатики и основывается на знаниях, полученных слушателями при изучении дисциплин «Дискретная математика», «Программирование», «Базы данных» и «Теория вероятностей и математическая статистика».

3. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения ОП магистратуры

Процесс изучения дисциплины направлен на формирование следующих элементов компетенций в соответствии с ФГОС ВО, ОП ВО и приобретения следующих знаний, умений, навыков и (или) опыта деятельности:

Формируемая компетенция (код и формулировка)	Индикатор достижения компетенции (код и формулировка)	Перечень планируемых результатов обучения
Общепрофессиональные компетенции		
ОПК-4 Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности.	ОПК-4.1. Знает основные идеи теории и практики применения информационно-коммуникационных технологий в области своей профессиональной деятельности; ОПК-4.2. Может применять технологии, а также самостоятельно расширять и углублять знания в области информационных технологий; ОПК-4.3. Способен решать задачи профессиональной деятельности с использованием информационно-коммуникационных технологий и с учетом основных требований информационной безопасности.	<u>Знать:</u> – алгоритмы построения деревьев решений и байесовской классификации, их применение для решения задачи классификации в прикладных задачах. – базовые алгоритмы кластеризации числовых и категориальных данных; – алгоритм Apriori поиска ассоциативных правил в базах транзакций; – принципы решения задач классификации с использованием искусственных нейронных сетей. <u>Уметь:</u> – формулировать задачи анализа данных разного типа; <u>Иметь навыки:</u> использования аналитической платформ Deductor Studio для решения прикладных задач обработки массивов данных алгоритмами Data Mining

		<u>Знать:</u>
		<p>– математические методы решения основных задач анализа БД больших размеров;</p> <p>– основы модели распределенных вычислений MapReduce.</p> <p><u>Уметь</u> Проводить препроцессинг данных (очистка данных, заполнение пропусков в БД); анализировать возможности аналитических платформ для решения конкретных задач обработки данных</p>

4. Объем, структура и содержание дисциплины

Общая трудоемкость дисциплины составляет 3 зач. ед., 108 акад. час.

№ п/п	Темы (разделы) дисциплины, их содержание	Семестр	Виды учебных занятий, включая самостоятельную работу студентов, и их трудоемкость (в академических часах)					Формы текущего контроля успеваемости	
			лекции	практические	лабораторные	консультации	аттестационные испытания	самостоятельная работа	Форма промежуточной аттестации (по семестрам)
			Контактная работа						
1.	Задачи интеллектуального анализа данных (DataMining).	3	2	2				14	Решение задач, опрос по теме
2.	Задачи классификации данных	3	2	2				14	Решение задач, опрос по теме
3.	Поиск ассоциативных правил. Секвенциальный анализ.	3	2	2				14	Решение задач, опрос по теме
4.	Задача кластеризации данных	3	2	2				14	Решение задач, опрос по теме
5.	Технологии Data Mining на базе платформы Deductor Studio	3	2	2		2		10	Решение задач на базе аналитической платформы Deductor Studio
6.	Концепции больших данных (Big Data)	3	7	7		2		8	Опрос по теме

	Всего за 3 семестр		17	17		4		70	
			17	17		4		70	Экзамен

Содержание разделов дисциплины:

РАЗДЕЛ 1. Задачи интеллектуального анализа данных (Data Mining)

Формы представления наборов данных Типы данных (векторные, категориальные, порядковые, неструктурированные).

Особенности данных, накопленных в компаниях. Формализация данных. Корреляционный анализ числовых и ранжированных данных.

Задачи интеллектуального анализа данных в маркетинговых и социологических исследованиях, прогнозирования, технической и медицинской диагностики

РАЗДЕЛ 2. Задачи классификации данных

Формальная постановка задачи классификации Алгоритмы классификации векторных данных (kNN – метод «к ближайших соседей», линейные классификаторы).

Классификация категориальных данных (деревья решений). Вероятностная классификация (байесовский классификатор). Нейросетевые алгоритмы классификации.

РАЗДЕЛ 3. Поиск ассоциативных правил. Секвенциальный анализ.

Базы транзакций, ассоциативные правила, показатели достоверности и поддержки ассоциативных правил. Алгоритм Apriori построения ассоциативных правил. Определение значимости и полезности ассоциативных правил, показатели их характеризующие. Постановка и решение задачи секвенциального анализа.

РАЗДЕЛ 4. Задача кластеризации данных.

Постановка задачи кластеризации. Графовые алгоритмы кластеризации. Алгоритмы k-means и «ближайшего соседа». Иерархическая кластеризация данных, основные подходы. Агломеративные и дивизионные методы. Метрика в пространстве кластеров. Кластеризация категориальных данных, алгоритм CLOPE.

РАЗДЕЛ 5. Технологии Data Mining на базе платформы Deductor Studio

Задачи анализа данных, решаемые в Deductor Studio. Основные компоненты Deductor Studio. Создание сценариев для типовых задач анализа данных (линейная регрессия, классификация, кластеризация, поиск ассоциативных правил)

РАЗДЕЛ 6. Концепции больших данных (Big Data).

Понятие и примеры больших данных. Базовые принципы обработки больших данных (горизонтальная масштабируемость и др.). Модель распределённых вычислений MapReduce и технология NoSQL.

5. Образовательные технологии, используемые при осуществлении образовательного процесса по дисциплине

В процессе обучения используются следующие образовательные технологии:

Вводная лекция – дает первое целостное представление о дисциплине и ориентирует студента в системе изучения данной дисциплины. Студенты знакомятся с назначением и задачами курса, его ролью и местом в системе учебных дисциплин и в системе подготовки в целом. Дается краткий обзор курса, история развития науки и практики, достижения в этой сфере, имена известных ученых, излагаются перспективные направления исследований. На этой лекции высказываются методические и организационные особенности работы в рамках данной дисциплины, а также дается анализ рекомендуемой учебно-методической литературы.

Академическая лекция (или лекция общего курса) – последовательное изложение материала, осуществляемое преимущественно в виде монолога преподавателя. Требования к академической лекции: современный научный уровень и насыщенная информативность, убедительная аргументация, доступная и понятная речь, четкая структура и логика, наличие ярких примеров, научных доказательств, обоснований, фактов.

Практическое занятие – занятие, посвященное освоению конкретных умений и навыков и закреплению полученных на лекции знаний.

На практических занятиях студенты решают поставленные перед ними задачи под руководством (контролем) преподавателя. Обсуждение процесса решения задачи и оценка правильности полученного результата (постановки задачи, выбора метода ее решения, проверка полученного результата и т.д.) в ходе практического занятия производится коллективно студентами под руководством преподавателя.

Консультации – групповые занятия, являющиеся одной из форм контроля самостоятельной работы студентов. На консультациях по просьбе студентов рассматриваются наиболее сложные моменты при освоении материала дисциплины, преподаватель отвечает на вопросы студентов, которые возникают у них в процессе самостоятельной работы, обсуждаются результаты решения заданий, выполненных студентами самостоятельно

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

В процессе осуществления образовательного процесса используются:

- для формирования текстов материалов для промежуточной и текущей аттестации, для разработки документов, презентаций, для работы с электронными таблицами -программы OfficeStd 2013 RUS OLP NL Acdmc 021-10232, LibreOffice (свободное), издательская система LaTeX;
- аналитическая платформа Deductor Studio Academic v.6.2.4 и выше (свободно распространяемая в Интернете фирмой BaseGroup Labs);
- для поиска учебной литературы библиотеки ЯрГУ– Автоматизированная библиотечная информационная система "БУКИ-NEXT" (АБИС "Буки-Next").

7. Перечень основной и дополнительной учебной литературы, ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины

а) основная:

1. Лесковец Ю. Анализ больших наборов данных / Ю. Лесковец, А. Раджараман, Дж Ульман – Пер. с англ. Слинкин А.А. – М.: ДМК Пресс, 2016.
2. Чубукова, И. А., Data Mining : учеб. пособие для вузов / И. А. Чубукова. - 2-е изд., испр., М., Интернет-Ун-т Информационных Технологий; БИНОМ. Ла, 2013, 382с

б) дополнительная:

3. Боровиков, В. П., Популярное введение в современный анализ данных в системе STATISTICA : методология и технология современного анализа данных : учеб. пособие для вузов / В. П. Боровиков, М., Горячая линия - Телеком, 2015, 288с

в) ресурсы сети «Интернет»

4. Электронные каталоги Научной библиотеки ЯрГУ им. П.Г. Демидова (http://www.lib.uni Yar.ac.ru/opac/bk_one_find.php)
5. Электронная картотека «Книгообеспеченность» Научной библиотеки ЯрГУ им. П.Г. Демидова (http://www.lib.uni Yar.ac.ru/opac/bk_bookreq_find.php)
6. Электронно-библиотечная система «Юрайт» (<https://urait.ru/>).
7. Электронно-библиотечная система «Лань» (<https://e.lanbook.com/>).

8. Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине

Материально-техническая база, необходимая для осуществления образовательного процесса по дисциплине включает в свой состав специальные помещения:

- учебные аудитории для проведения занятий лекционного типа;
- учебные аудитории для проведения групповых и индивидуальных консультаций,
- учебные аудитории для проведения текущего контроля и промежуточной аттестации;
- помещения для самостоятельной работы;
- помещения для хранения и профилактического обслуживания технических средств обучения.

Специальные помещения укомплектованы средствами обучения, служащими для представления учебной информации большой аудитории.

Для проведения занятий лекционного типа предлагаются наборы демонстрационного оборудования и учебно-наглядных пособий, хранящиеся на электронных носителях и обеспечивающие тематические иллюстрации, соответствующие рабочим программам дисциплин.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду организации.

Число посадочных мест в лекционной аудитории больше либо равно списочному составу потока, а в аудитории для практических занятий (семинаров) – списочному составу группы обучающихся.

Автор(ы) :

доцент кафедры компьютерных сетей, к.т.н А.А. Короткин

**Приложение №1 к рабочей программе дисциплины
«Технологии больших данных и Data Mining»**

**Фонд оценочных средств
для проведения текущей и промежуточной аттестации студентов
по дисциплине**

1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

1.1. Контрольные задания и иные материалы, используемые в процессе текущей аттестации

Задания для самостоятельной работы

Задания по теме №2 (компетенция ПК-11)

Нейросетевые алгоритмы классификации

Задания по теме № 3 (компетенция ПК-11)

Определение значимости и полезности ассоциативных правил, показатели их характеризующие.

Задания по теме № 4 (компетенция ПК-11)

Кластеризация категориальных данных, алгоритм CLOPE.

Задания по теме № 5 (компетенция ПК-4)

Создание сценария в Deductor Studio для решения задачи поиска ассоциативных правил.

Задания по теме № 6 (компетенция ПК-4)

Модель распределённых вычислений MapReduce и технологии NoSQL

Описание процедуры выставления оценивания сформированности компетенций

Результат определяется оценками «отлично» (высокий уровень), «хорошо» (продвинутый уровень), «удовлетворительно» (пороговый уровень), «неудовлетворительно» (компетенция не сформирована). Решение заданий свидетельствует об уровне сформированности компетенции ПК-11 (задания по темам 2, 3, 4) и компетенции ПК-4 (задания по темам 5, 6).

Для оценивания ответов (по каждой из компетенций):

Оценка «отлично»:

- Все задания решены верно,
- Оформлены по требованиям,
- Решение изложено достаточно полно и чётко.
- Даны правильные формулировки, точные определения, понятия терминов.

Оценка «хорошо»:

- Все задания решены верно,
- Оформлены по требованиям,
- Но, решение изложено недостаточно полно и чётко (не менее 70 % от полного)
- При изложении были допущены 1-2 несущественные ошибки;
- Даны правильные формулировки, точные определения, понятия терминов.

Оценка «удовлетворительно»:

- Более половины заданий решены верно,

- Все задания оформлены по требованиям,
- Решение изложено недостаточно полно и чётко (не менее 70 % от полного), при изложении некоторых заданий допущена 1 существенная ошибка, приводящая к неверному ответу.
- Студент знает и понимает основные положения данной темы, но допускает неточности в формулировки понятий.

Оценка «неудовлетворительно»:

- Более половины заданий решены неверно,
- Решение изложено неполно и нечётко (менее 50 % от полного), при изложении многих задач были допущены существенные ошибки, приводящая к неверному ответу.

Список вопросов к зачету

Компетенция ПК-11:

1. Коэффициенты корреляции числовых и ранжированных данных
2. Уравнение линейной регрессии.
3. Формальная постановка задачи классификации на основе обучающей выборки. Алгоритм классификации kNN.
4. Деревья решений. Пример построения дерева решений (алгоритм CART).
5. Алгоритм вероятностной классификации по методу Байеса. Решение простой задачи медицинской диагностики методом Байеса.
6. Нейросетевые алгоритмы классификации

Компетенция ПК-4:

7. Формальная постановка задачи кластеризации данных. Описание алгоритма k-means.
8. Алгоритм кластеризации категориальных данных (транзакций)
9. Задача иерархической кластеризации.
10. Классификация текстовых данных. Метод Байеса для классификации текстов.
11. Ассоциативные правила, их характеристики. Алгоритм Apriori построения ассоциативных правил.
12. Модель распределенных вычислений MapReduce

Критерии оценивания зачета

Фактически оценка зачета выполняется на основе заданий по темам 2–6. Если по итогу их выполнения обе компетенции сформированы не ниже чем на пороговом уровне, выставляется оценка «зачтено».

Если одна из компетенций сформирована не ниже чем на пороговом уровне, а другая не сформирована, проводится устная беседа по вопросам из тематики несформированной компетенции (список вопросов известен студенту заранее). Если в ходе беседы студент продемонстрировал достаточное знание материала (хотя бы на пороговом уровне), ему также выставляется оценка «зачтено». В противном случае выставляется оценка «не зачтено»

Если обе компетенции не сформированы, выставляется оценка «не зачтено».

Критерии выставления оценок «зачтено» и «не зачтено» приведены в следующей таблице.

Показатели	Критерии	Оценка
------------	----------	--------

<p>ПК-11 <u>Знать:</u> – алгоритмы построения деревьев решений и байесовской классификации, их применение для решения задачи классификации в прикладных задачах. – базовые алгоритмы кластеризации числовых и категориальных данных; – алгоритм Apriori поиска ассоциативных правил в базах транзакций; – принципы решения задач классификации с использованием искусственных нейронных сетей.</p>	<p>ПК-11 Студентом дан полный, в логической последовательности развернутый ответ на поставленный вопрос, где он продемонстрировал знания предмета в полном объеме учебной программы, достаточно глубоко осмысливает дисциплину. ПК-4 Студент самостоятельно, и исчерпывающе отвечает на дополнительные вопросы, приводит собственные примеры по проблематике поставленного вопроса, решил предложенные практические задания без ошибок.</p>	<p>«Зачтено»</p>
<p><u>Уметь:</u> – формулировать задачи анализа данных разного типа; <u>Иметь навыки:</u> использования аналитической платформ Deductor Studio для решения прикладных задач обработки массивов данных алгоритмами Data Mining ПК-4 <u>Знать:</u> – математические методы решения основных задач анализа БД больших размеров; – основы модели распределенных вычислений MapReduce. <u>Уметь</u> Проводить препроцессинг данных (очистка данных, заполнение пропусков в БД); анализировать возможности аналитических платформ для решения конкретных задач обработки данных</p>	<p>ПК-11 Студентом дан ответ, который содержит ряд серьезных неточностей, обнаруживающий незнание процессов изучаемой предметной области, отличающийся неглубоким раскрытием темы, незнанием основных вопросов теории, несформированными навыками анализа явлений, процессов, неумением давать аргументированные ответы, слабым владением монологической речью, отсутствием логичности и последовательности. ПК-4 Выводы поверхностны. Решение практических заданий не выполнено. Т.е студент не способен ответить на вопросы даже при дополнительных наводящих вопросах преподавателя.</p>	<p>«Не зачтено» (уровень не сформирован)</p>

Приложение 2. Перечень компетенций, этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкалы оценивания

2.1. Шкала оценивания сформированности компетенций и ее описание

Оценивание уровня сформированности компетенций в процессе освоения дисциплины осуществляется по следующей трехуровневой шкале:

Пороговый уровень - предполагает отражение тех ожидаемых результатов, которые определяют минимальный набор знаний и (или) умений и (или) навыков, полученных студентом в результате освоения дисциплины. Пороговый уровень является обязательным уровнем для студента к моменту завершения им освоения данной дисциплины.

Продвинутый уровень - предполагает способность студента использовать знания, умения, навыки и (или) опыт деятельности, полученные при освоении дисциплины, для решения профессиональных задач. Продвинутый уровень превосходит пороговый уровень по нескольким существенным признакам.

Высокий уровень - предполагает способность студента использовать потенциал интегрированных знаний, умений, навыков и (или) опыта деятельности, полученных при освоении дисциплины, для творческого решения профессиональных задач и самостоятельного поиска новых подходов в их решении путем комбинирования и использования известных способов решения применительно к конкретным условиям. Высокий уровень превосходит пороговый уровень по всем существенным признакам.

2.2. Перечень компетенций, этапы их формирования, описание показателей и критериев оценивания компетенций на различных этапах их формирования

Код компетенции	Форма контроля	Этапы формирования (№ темы (раздела))	Показатели оценивания	Шкала и критерии оценивания компетенций на различных этапах их формирования		
				Пороговый уровень	Продвинутый уровень	Высокий уровень
Профессиональные компетенции						
ПК-11	Самостоятельные работы по темам Зачет	Разделы №№ 1-5	<u>Знать:</u> постановки базовых задач Data Mining и алгоритмы их решения. <u>Уметь:</u> применять алгоритмы Data Mining для решения прикладных задач аналитического анализа данных <u>Владеть навыками:</u> решения практических задач на аналитической платформе Deductor	Знать постановку задач классификации и кластеризации данных, поиска ассоциативных правил, области их применения	Знать (помимо указанных в разделе Пороговый уровень) основные алгоритмы технологии Data Mining	Дополнительно к знаниям и умениям, указанных в разделе Продвинутый уровень, Знать - концепции больших данных (Big Data), модель распределённых вычислений MapReduce для параллельных вычислений
ПК-4	Самостоятельные работы по темам Зачет.	Разделы №№ 1-6	<u>Знать:</u> области применения инструментов Data Mining для анализа данных; <u>Уметь:</u> адаптировать известные алгоритмы для решения нестандартных задач обработки данных.	Знать базовые алгоритмы решения задач классификации и кластеризации данных, поиска ассоциативных правил: kNN, метод Байеса, деревья решений, алгоритм кластеризации k-means, алгоритм Apriori.	Знать (помимо указанного в разделе Пороговый уровень) - алгоритмы кластеризации категориальных данных, - нейросетевые алгоритмы решения задач прогнозирования и классификации. Уметь решать задачи классификации векторных	Дополнительно к знаниям и умениям, указанных в разделе Продвинутый уровень, знать - алгоритмы секвенциального анализа временных последовательностей событий. уметь - оценивать эффективность

					данных на аналитической платформе Deductor	различных алгоритмов решения задач анализа эмпирических данных
--	--	--	--	--	--	--

3. Методические рекомендации преподавателю по процедуре оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций

Целью процедуры оценивания является определение степени овладения студентом ожидаемыми результатами обучения (знаниями, умениями, навыками и (или) опытом деятельности).

Процедура оценивания степени овладения студентом ожидаемыми результатами обучения осуществляется с помощью методических материалов, представленных в разделе «Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций»

3.1 Критерии оценивания степени овладения знаниями, умениями, навыками и (или) опытом деятельности, определяющие уровни сформированности компетенций

Пороговый уровень (общие характеристики):

- владение основным объемом знаний по программе дисциплины;
- знание основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы без существенных ошибок;
- владение инструментарием дисциплины, умение его использовать в решении стандартных (типовых) задач;
- способность самостоятельно применять типовые решения в рамках рабочей программы дисциплины;
- усвоение основной литературы, рекомендованной рабочей программой дисциплины;
- знание базовых теорий, концепций и направлений по изучаемой дисциплине;
- самостоятельная работа на практических и лабораторных занятиях, периодическое участие в групповых обсуждениях, достаточный уровень культуры исполнения заданий.

Продвинутый уровень (общие характеристики):

- достаточно полные и систематизированные знания в объеме программы дисциплины;
- использование основной терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать выводы;
- владение инструментарием дисциплины, умение его использовать в решении учебных и профессиональных задач;
- способность самостоятельно решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;
- усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в базовых теориях, концепциях и направлениях по изучаемой дисциплине и давать им сравнительную оценку;
- самостоятельная работа на практических и лабораторных занятиях, участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

Высокий уровень (общие характеристики):

- систематизированные, глубокие и полные знания по всем разделам дисциплины;
- точное использование терминологии данной области знаний, стилистически грамотное, логически правильное изложение ответа на вопросы, умение делать обоснованные выводы;
- безупречное владение инструментарием дисциплины, умение его использовать в постановке и решении научных и профессиональных задач;
- способность самостоятельно и творчески решать сложные задачи (проблемы) в рамках рабочей программы дисциплины;

- полное и глубокое усвоение основной и дополнительной литературы, рекомендованной рабочей программой дисциплины;
- умение ориентироваться в основных теориях, концепциях и направлениях по изучаемой дисциплине и давать им критическую оценку;
- активная самостоятельная работа на практических и лабораторных занятиях, творческое участие в групповых обсуждениях, высокий уровень культуры исполнения заданий.

3.2 Описание процедуры выставления оценки

В зависимости от уровня сформированности каждой компетенции по окончании освоения дисциплины студенту выставляется оценка. Для дисциплин, изучаемых в течение нескольких семестров, оценка может выставляться не только по окончании ее освоения, но и в промежуточных семестрах. Вид оценки («отлично», «хорошо», «удовлетворительно», «неудовлетворительно», «зачтено», «незачтено») определяется рабочей программой дисциплины в соответствии с учебным планом.

Оценка «отлично» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована на высоком уровне.

Оценка «хорошо» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на продвинутом уровне.

Оценка «удовлетворительно» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на пороговом уровне.

Оценка «неудовлетворительно» выставляется студенту, у которого хотя бы одна компетенция (полностью или частично формируемая данной дисциплиной) сформирована ниже, чем на пороговом уровне.

Оценка «зачет» выставляется студенту, у которого каждая компетенция (полностью или частично формируемая данной дисциплиной) сформирована не ниже, чем на пороговом уровне.

Оценка «незачтено» выставляется студенту, у которого хотя бы одна компетенция (полностью или частично формируемая данной дисциплиной) сформирована ниже, чем на пороговом уровне.

Приложение №2 к рабочей программе дисциплины «Технологии больших данных и Data Mining»

Методические указания для студентов по освоению дисциплины

Основной формой изложения учебного материала по дисциплине «Технологии больших данных и Data Mining» являются лекции, причем в достаточно большом объеме. Это связано с тем, что в основе указанных технологий лежит особый математический аппарат, с помощью которого решаются довольно сложные и громоздкие задачи. По большому числу тем предусмотрены практические занятия, на которых происходит закрепление лекционного материала путем применения его к конкретным задачам и отработка практических навыков программирования.

Для успешного освоения дисциплины очень важно решение достаточно большого количества задач, как в аудитории, так и самостоятельно в качестве домашних заданий. Примеры решения задач разбираются на лекциях и практических занятиях, при необходимости по наиболее трудным темам проводятся дополнительные консультации. Основная цель решения задач – помочь усвоить фундаментальные понятия и основы технологии Big Data и Data Mining. Для решения всех задач необходимо знать и понимать лекционный материал. Поэтому в процессе изучения дисциплины рекомендуется регулярное повторение пройденного лекционного материала. Материал, законспектированный на лекциях, необходимо дома еще раз прорабатывать и при необходимости дополнять информацией, полученной на консультациях, практических занятиях или из учебной литературы.

Большое внимание должно быть уделено выполнению домашней работы. В качестве заданий для самостоятельной работы дома студентам предлагаются задачи, аналогичные разобранным на лекциях и практических занятиях или немного более сложные, которые являются результатом объединения нескольких базовых задач.

Для проверки и контроля усвоения теоретического материала, приобретенных практических навыков работы с аппаратом технологий Data Mining, в течение обучения проводятся мероприятия текущей аттестации в виде опроса по темам самостоятельных работ. Также проводятся консультации (при необходимости) по разбору заданий для самостоятельной работы, которые вызвали затруднения.

В конце семестра изучения дисциплины студенты сдают зачет, Зачет по итогам первого семестра выставляется по итогам тестирования и краткого собеседования по его результатам.

Освоить вопросы, излагаемые в процессе изучения дисциплины «Технологии больших данных и Data Mining» самостоятельно студенту крайне сложно. Это связано со сложностью изучаемого материала и большим объемом курса. Поэтому посещение всех аудиторных занятий является совершенно необходимым. Без упорных и регулярных занятий в течение семестра сдать зачет по итогам изучения дисциплины студенту практически невозможно.

Учебно-методическое обеспечение самостоятельной работы студентов по дисциплине

Для самостоятельной работы особенно рекомендуется использовать учебную литературу. Также для подбора учебной литературы рекомендуется использовать широкий спектр интернет-ресурсов:

1. Электронно-библиотечная система «Университетская библиотека online» (www.biblioclub.ru) - электронная библиотека, обеспечивающая доступ к наиболее востребованным материалам-первоисточникам, учебной, научной и художественной литературе ведущих издательств (регистрация в электронной библиотеке – только в сети университета. После регистрации работа с системой возможна с любой точки доступа в Internet.).

2. Информационная система "Единое окно доступа к образовательным ресурсам" (<http://window.edu.ru/library>).

Целью создания информационной системы "Единое окно доступа к образовательным ресурсам" (ИС "Единое окно ") является обеспечение свободного доступа к интегральному каталогу образовательных интернет-ресурсов и к электронной библиотеке учебно-методических материалов для общего и профессионального образования.

Информационная система "Единое окно доступа к образовательным ресурсам" создана по заказу Федерального агентства по образованию в 2005-2008 гг. Главной разработчик проекта - Федеральное государственное автономное учреждение Государственный научно-исследовательский институт информационных технологий и телекоммуникаций (ФГАУ ГНИИ ИТТ "Информика") www.informika.ru.

ИС "Единое окно" объединяет в единое информационное пространство электронные ресурсы свободного доступа для всех уровней образования в России. Разделы этой системы:

- Электронная библиотека является крупнейшим в российском сегменте Интернета хранилищем полнотекстовых версий учебных, учебно-методических и научных материалов с открытым доступом. Библиотека содержит более 30 000 материалов, источниками которых являются более трехсот российских вузов и других образовательных и научных учреждений. Основу наполнения библиотеки составляют электронные версии учебно-методических материалов, подготовленные в вузах, прошедшие рецензирование и рекомендованные к использованию советами факультетов, учебно-методическими комиссиями и другими вузовскими структурами, осуществляющими контроль учебно-методической деятельности.

- Интегральный каталог образовательных интернет-ресурсов содержит представленные в стандартизированной форме метаданные внешних ресурсов, а также содержит описания полнотекстовых публикаций электронной библиотеки. Общий объем каталога превышает 56 000 метаописаний (из них около 25 000 - внешние ресурсы). Расширенный поиск в "Каталоге" осуществляется по названию, автору, аннотации, ключевым словам с возможной фильтрацией по тематике, предмету, типу материала, уровню образования и аудитории.

- Избранное. В разделе представлены подборки наиболее содержательных и полезных, по мнению редакции, Интернет-ресурсов для общего и профессионального образования.

- Библиотеки вузов. Раздел содержит подборки сайтов вузовских библиотек, электронных каталогов библиотек вузов и полнотекстовых электронных библиотек вузов.

Для самостоятельного подбора литературы в библиотеке ЯрГУ рекомендуется использовать:

1. Личный кабинет (http://lib.uniyar.ac.ru/opac/bk_login.php) дает возможность получения on-line доступа к списку выданной в автоматизированном режиме литературы, просмотра и копирования электронных версий изданий сотрудников университета (учеб.и метод. пособия, тексты лекций и т.д.) Для работы в «Личном кабинете» необходимо зайти на сайт

Научной библиотеки ЯрГУ с любой точки, имеющей доступ в Internet, в пункт меню «Электронный каталог»; пройти процедуру авторизации, выбрав вкладку «Авторизация», и заполнить представленные поля информации.

2. Электронная библиотека учебных материалов ЯрГУ

(http://www.lib.uniyaг.ac.ru/opac/bk_cat_find.php) содержит более 2500 полных текстов учебных и учебно-методических материалов по основным изучаемым дисциплинам, изданных в университете. Доступ в сети университета, либо по логину/пароллю.

3. Отметим три ЭБС, которые содержат большое количество полезных электронных ресурсов по дисциплине:

а) Электронная библиотечная система издательства «Юрайт» – это виртуальный читальный зал учебников и учебных пособий от авторов из ведущих вузов России по естественно-научным направлениям и специальностям. На сегодняшний день портфель издательства включает в себя более 3000 наименований учебной литературы. *Для ЯрГУ им. П. Г. Демидова открыт полнотекстовый доступ ко всем книгам с возможностью цитирования и создания закладок.* Работать с ресурсом можно из сети университета или удаленно, предварительно зарегистрировав свой личный кабинет, находясь внутри сети вуза.

б) Электронно-библиотечная система издательства "Лань" – это ресурс, предоставляющий online доступ к научным журналам и полнотекстовым коллекциям книг. Работать с ресурсом можно из сети вуза без предварительной регистрации или из любой точки мира, где есть доступ к сети "Интернет", предварительно зарегистрировав свой личный кабинет, находясь внутри сети вуза

в) Электронно-библиотечная система IPRbooks - уникальный ресурс, объединяющий лицензионную учебную и научную литературу, периодические издания, аудиокниги, видеокурсы, он-лайн тесты. IPRbooks содержит более 127 000 изданий, из которых более 40 000 - учебные и научные издания по различным дисциплинам, около 700 наименований российских и зарубежных журналов, более 2000 аудиоизданий.

Работать с ресурсом можно из сети университета или удаленно, после авторизации, для которой новым пользователям нужно получить логин и пароль в библиотеке (e-mail eresurs@uniyaг.ac.ru) . После этого необходимо пройти личную регистрацию и в дальнейшем входить в ЭБС под своими учетными данными

4. Электронная картотека «Книгообеспеченность»

(http://www.lib.uniyaг.ac.ru/opac/bk_bookreq_find.php) раскрывает учебный фонд научной библиотеки ЯрГУ, предоставляет оперативную информацию о состоянии книгообеспеченности дисциплин основной и дополнительной литературой, а также цикла дисциплин и специальностей. Электронная картотека «Книгообеспеченность» доступна в сети университета и через Личный кабинет.