

Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ ФГБОУ ВО «КемГУ»
Дата и время: 2024-02-21 00:00:00
471086fad29a3b30e244c728abc3661ab35c9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Кузбасский гуманитарно-педагогический институт
федерального государственного бюджетного образовательного учреждения
высшего образования
«Кемеровский государственный университет»

Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан А.В. Фомина
«09» февраля 2023 г.

**Рабочая программа дисциплины
К.М.07.01 Машинное обучение**

Направление подготовки
01.03.02 Прикладная математика и информатика

Направленность (профиль) подготовки
ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ ДАННЫХ

Программа бакалавриата

Квалификация выпускника
бакалавр

Форма обучения
Очная

Год набора 2022

Новокузнецк 2023

Оглавление

1	Цель дисциплины.	3
1.1	Формируемые компетенции	3
1.2	Индикаторы достижения компетенций	3
1.3	Знания, умения, навыки (ЗУВ) по дисциплине	3
2	Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.	4
3.	Учебно-тематический план и содержание дисциплины.	4
3.1	Учебно-тематический план	4
3.2.	Содержание занятий по видам учебной работы	5
4	Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.	6
5	Материально-техническое, программное и учебно-методическое обеспечение дисциплины.	6
5.1	Учебная литература	6
5.2	Материально-техническое и программное обеспечение дисциплины.	7
5.3	Современные профессиональные базы данных и информационные справочные системы.	7
6	Иные сведения и (или) материалы.	8
6.1.	Примерные вопросы и задания / задачи для промежуточной аттестации .	8

1 Цель дисциплины.

В результате освоения данной дисциплины у обучающегося должны быть сформированы компетенции основной профессиональной образовательной программы магистратуры (далее - ОПОП): ПК-1

Содержание компетенций как планируемых результатов обучения по дисциплине см. таблицы 1 и 2.

1.1 Формируемые компетенции

Таблица 1 - Формируемые дисциплиной компетенции

Наименование вида компетенции (универсальная, общепрофессиональная, профессиональная)	Наименование категории (группы) компетенций	Код и название компетенции
профессиональная		ПК-1 Способен проводить аналитические исследования с применением технологий больших данных

1.2 Индикаторы достижения компетенций

Таблица 2 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	Дисциплины и практики, формирующие компетенцию ОПОП
ПК-1 Способен проводить аналитические исследования с применением технологий больших данных	ПК 1.1 Способен осуществлять выявление, формирование и согласование требований к результатам аналитических работ с применением технологий больших данных ПК 1.2 Способен планировать и организовывать аналитические работы с использованием технологий больших данных ПК 1.3 Способен подготавливать данные для проведения аналитических работ по исследованию больших данных ПК 1.4 Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика	К.М.07.01 Машинное обучение К.М.07.02 Аналитика данных К.М.07.03 Технологии работы с большими данными К.М.07.04 Технологии работы с открытыми данными К.М.07.ДВ.01.01 Машинное обучение с подкреплением/ К.М.07.ДВ.01.02 Глубокое обучение К.М.09.06(Пд) Преддипломная практика

1.3 Знания, умения, навыки (ЗУВ) по дисциплине

Таблица 3 – Знания, умения, навыки, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
----------------------------	--	---

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
ПК-1 Способен проводить аналитические исследования с применением технологий больших данных	ПК 1.4 Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика	<p>Знать:</p> <ul style="list-style-type: none"> – современные методы, используемые в машинном обучении – тенденции развития, научные и прикладные достижения в области машинного обучения <p>Уметь:</p> <ul style="list-style-type: none"> – анализировать и модифицировать методы машинного обучения для определения наиболее подходящего при прогнозе конкретного процесса – использовать методы машинного обучения для решения научно-исследовательских и прикладных задач. <p>Владеть:</p> <ul style="list-style-type: none"> – навыками использования современных методов машинного обучения для решения прикладных задач

2 Объём и трудоёмкость дисциплины по видам учебных занятий. Формы промежуточной аттестации.

Таблица 4 – Объем и трудоемкость дисциплины по видам учебных занятий

Общая трудоемкость и виды учебной работы по дисциплине, проводимые в разных формах	Объём часов по формам обучения
	ОФО
1 Общая трудоемкость дисциплины	108
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	51
Аудиторная работа (всего):	48
в том числе:	
лекции	16
лабораторные работы	32
Внеаудиторная работа (всего):	3
в том числе, индивидуальная работа обучающихся с преподавателем	
подготовка курсовой работы /контактная работа	3
3 Самостоятельная работа обучающихся (всего)	57
4 Промежуточная аттестация обучающегося и объём часов, выделенный на промежуточную аттестацию: - зачет 6 семестр	

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 5 - Учебно-тематический план очной формы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоемкость (всего час.)	Трудоемкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			ОФО		СРС	
			Аудиторн. занятия	лекц.		
1-8	1. Введение в методы машинного обучения	20	4	8	8	Отчет по лабораторным работам
9-	2. Метрические методы машинного	21	4	8	9	Отчет по лабораторным

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость (всего час.)	Трудоёмкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			ОФО			
			Аудиторн. занятия		СРС	
			лекц.	лаб.		
16	обучения					работам
1-8	3. Линейные методы машинного обучения	32	4	8	20	Отчет по лабораторным работам
9-16	4. Кластеризация и вероятностное моделирование данных	32	4	8	20	Отчет по лабораторным работам
	Промежуточная аттестация	3				Курсовая работа, зачет 6 семестр
	Всего:	108	16	32	57	3

3.2. Содержание занятий по видам учебной работы

Таблица 6 – Содержание дисциплины

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
<i>Содержание лекционного курса</i>		
1	Введение в методы машинного обучения	Общая постановка задачи машинного обучения. Обучение с учителем и без учителя, обучение с подкреплением. Задачи классификации, восстановления регрессии, предсказания. Модели алгоритмов. Признаки. Типы признаков. Понятие функционала качества. Вероятностная постановка задачи. Оценка обобщающей способности. Проблема переобучения. Критерии оценки качества работы алгоритмов машинного обучения. ROC-кривые. Примеры практических задач машинного обучения
2	Метрические методы машинного обучения	Обобщенный метрический классификатор. Виды метрик. Метод ближайшего соседа. Алгоритм k-ближайших соседей. Взвешенная версия алгоритма k-ближайших соседей. Метод окна Парзена. Метод потенциальных функций. Понятие эталона. Отступы и классификация объектов.
3	Линейные методы машинного обучения	Постановка задач линейной регрессии и линейной классификации. Метод наименьших квадратов в матричной форме. Аналитическое решение. Регуляризация в задач регрессии. Мультиколлинеарность и плохая обусловленность ковариационной матрицы. Гребневая регрессия. Метод лассо. Линейные классификаторы. Метод стохастического градиента. Улучшение сходимости метода SGD. Логистическая регрессия. Метод опорных векторов. Линейно разделимые выборки. Двойственная задача. Нелинейные обобщения. Возможные виды ядер
4	Кластеризация и вероятностное моделирование данных	Методы кластеризации. Типы кластерных структур. Функционал качества кластеризации. EM-алгоритм. Метод k-средних. Иерархическая кластеризация. Формула Ланса-Уильямса. Быстрая агрегативная кластеризация.
<i>Содержание лабораторных занятий</i>		
1.	Введение в методы машинного обучения	Основы Python. Работа с библиотекой NumPy Первичный анализ данных с Pandas
2	Метрические методы машинного обучения	Визуальный анализ данных
3	Линейные методы машинного обучения	Деревья принятия решений Линейные методы Линейная регрессия, Lasso и RF-регрессия. Построение и отбор признаков
4	Кластеризация и вероятностное моделирование данных	Обучение без учителя. Метод главных компонент. Кластеризация

4 Порядок оценивания успеваемости и сформированности компетенций обучающегося в текущей и промежуточной аттестации.

Для положительной оценки по результатам освоения дисциплины обучающемуся необходимо выполнить все установленные виды учебной работы. Оценка результатов работы обучающегося в баллах (по видам) приведена в таблице 7.

Таблица 7 - Шкала и показатели оценивания результатов учебной работы обучающихся по видам в балльно-рейтинговой системе (БРС)

6 семестр

Учебная работа (виды)	Сумма баллов	Виды и результаты учебной работы	Оценка в аттестации (шкала и показатели оценивания)	Баллы
Текущая учебная работа в семестре (Посещение занятий по расписанию и выполнение заданий)	80	Лекционные занятия (конспект). (8 занятий)	3 баллов посещение и конспект 1 лекционного занятия	15-24
		Курсовая работа (приведенные баллы)	8 баллов (пороговое значение) 16 баллов (максимальное значение)	8-16
		Лабораторные работы	Отчет по лабораторной работе (4 отчета): 5 баллов (выполнено 51 - 65% заданий) 7 балла (выполнено 66 - 85% заданий) 10 баллов (выполнено 86 - 100% заданий)	18-40
Итого по текущей работе в семестре				41 - 80
Промежуточная аттестация (зачет)	20	Решение задачи 1.	5 баллов (пороговое значение) 10 баллов (максимальное значение)	5-10
		Решение задачи 2.	5 баллов (пороговое значение) 10 баллов (максимальное значение)	5-10
Итого по промежуточной аттестации (зачет)				10-20
Суммарная оценка по дисциплине: Сумма баллов текущей и промежуточной аттестации				51 – 100 б.

В промежуточной аттестации оценка выставляется в ведомость в 100-балльной шкале и в буквенном эквиваленте (таблица 8)

Таблица 8 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных баллов	Уровни освоения дисциплины и компетенций	Экзамен		Зачет
		Оценка	Буквенный эквивалент	Буквенный эквивалент
86 - 100	Продвинутый	5	отлично	Зачтено
66 - 85	Повышенный	4	хорошо	
51 - 65	Пороговый	3	удовлетворительно	
0 - 50	Первый	2	неудовлетворительно	Не зачтено

5 Материально-техническое, программное и учебно-методическое обеспечение дисциплины.

5.1 Учебная литература

Основная учебная литература

1. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных / П. Флах. — Москва : ДМК Пресс, 2015. — 400 с. — ISBN 978-5-97060-273-7. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/69955>. — Режим доступа: для авториз. пользователей.

Дополнительная учебная литература

1. Гладилин, П. Е. Технологии машинного обучения : учебно-методическое пособие / П. Е. Гладилин, К. О. Боченина. — Санкт-Петербург : НИУ ИТМО, 2020. — 75 с. — Текст :

электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/190885>. — Режим доступа: для авториз. пользователей.

2. Шалев-Шварц, Ш. Идеи машинного обучения : учебное пособие / Ш. Шалев-Шварц, Бен-ДавидШ. ; перевод с английского А. А. Слинкина. — Москва : ДМК Пресс, 2019. — 436 с. — ISBN 978-5-97060-673-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131686>. — Режим доступа: для авториз. пользователей.

3. Баймуратов, И. Р. Методы автоматизации машинного обучения : учебное пособие / И. Р. Баймуратов. — Санкт-Петербург : НИУ ИТМО, 2020. — 40 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/190871>. — Режим доступа: для авториз. пользователей.

4. Маккинни, У. Python и анализ данных / У. Маккинни ; перевод с английского А. А. Слинкина. — 2-ое изд., испр. и доп. — Москва : ДМК Пресс, 2020. — 540 с. — ISBN 978-5-97060-590-5. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/131721>. — Режим доступа: для авториз. пользователей.

5.2 Материально-техническое и программное обеспечение

дисциплины.

Учебные занятия по дисциплине проводятся в учебных аудиториях КГПИ ФГБОУ ВО КемГУ:

<p>610 Учебная аудитория (мультимедийная) для проведения:</p> <ul style="list-style-type: none"> - занятий лекционного типа; - текущего контроля и промежуточной аттестации. <p>Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья.</p> <p>Оборудование для презентации учебного материала: стационарное - компьютер, экран, проектор.</p> <p>Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), FoxitReader (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО).</p> <p>Интернет с обеспечением доступа в ЭИОС.</p>	<p>Учебный корпус №4.</p> <p>654079, Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19</p>
<p>502 Компьютерный класс.</p> <p>Учебная аудитория (мультимедийная) для проведения:</p> <ul style="list-style-type: none"> - занятий лекционного типа; - занятий семинарского (практического) типа; - занятий лабораторного типа; - групповых и индивидуальных консультаций; - самостоятельной работы; - текущего контроля и промежуточной аттестации. <p>Специализированная (учебная) мебель: доска меловая, столы компьютерные, стулья.</p> <p>Оборудование для презентации учебного материала: стационарное - компьютер, экран, проектор, наушники.</p> <p>Лабораторное оборудование: стационарное – компьютеры для обучающихся (16 шт.).</p> <p>Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), LibreOffice (свободно распространяемое ПО), Firefox 14 (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО), MicrosoftVisualStudio (MicrosoftImaginePremium 3 year по лицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021 г.), Среда статистических вычислений Rv.4.0.2 (свободно распространяемое ПО).</p> <p>Интернет с обеспечением доступа в ЭИОС.</p>	<p>Учебный корпус №4.</p> <p>654079, Кемеровская область, г. Новокузнецк, пр-кт Metallургов, д. 19</p>

5.3 Современные профессиональные базы данных и информационные справочные системы.

Перечень СПБД и ИСС по дисциплине

1. Информационная система «Общероссийский математический портал», режим доступа : <http://www.mathnet.ru/>

2. Информационная система «Экспонента» - центр инженерных технологий и моделирования, режим доступа :<http://www.exponenta.ru>
3. База данных Science Direct (более 1500 журналов издательства Elsevier, среди них издания по математике и информатике), режим доступа :<https://www.sciencedirect.com>
4. UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения - <http://archive.ics.uci.edu/ml/>
5. IAPR Education Committee & Resources — коллекция ссылок на образовательные ресурсы по распознаванию образов, машинному обучению, обработке сигналов, обработке изображений и компьютерному зрению, поддерживаемая Международной ассоциацией распознавания образов - <http://homepages.inf.ed.ac.uk/rbf/IAPR/index.php>
6. Портал по интеллектуальному анализу данных, поддерживаемый Григорием Пятецким-Шапиро - <http://www.kdnuggets.com/>
7. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - <http://machinelearning.ru>

6 Другие сведения и (или) материалы.

6.1. Примерные темы для курсовой работы

1. Распознавание изображений методами машинного обучения.
2. Прогнозирование исходов футбольных матчей с помощью машинного обучения.
3. Прогнозирование стихийных бедствий с помощью методов машинного обучения.
4. Классификация клиентов компании методами машинного обучения.
5. Разработка модели прогнозирования эффективности контекстной рекламы.
6. Разработка персонализированного агрегатора новостей.
7. Анализа финансовых рынков методами машинного обучения.
8. Разработка системы авторизации и аутентификации пользователей информационных систем на основе решения задач распознавания биометрических данных пользователя.
9. Автоматизированная информационная система распознавания символов рукописного ввода

6.2. Примерные вопросы и задания / задачи для промежуточной аттестации

Таблица 9 - Примерные теоретические вопросы и практические задания / задачи к зачету

Разделы и темы	Примерные теоретические вопросы	Примерные практические задания / задачи
Введение в методы машинного обучения	<ol style="list-style-type: none"> 1. Постановка задачи машинного обучения. Обучение с учителем и без учителя. Типы признаков 2. Типы задач. Линейные модели 3. Примеры задач обучения без учителя (3-5 примеров) 4. Примеры задач обучения с учителем (3-5 примеров) 	Покажите, что с ростом размерности пространства признаков при равномерном распределении точек в кубе $[0; 1]^d$ вероятность попасть в куб $[0; 0, 99]^d$ стремится к нулю. Это одна из иллюстраций проклятия размерностей (dimension curse).

	<p>5. Функционал качества. Минимизация эмпирического риска</p> <p>6. Переобучение. Методы валидации моделей. Кросс-валидация</p>	<p>Попробуйте придумать или найти еще какую-нибудь иллюстрацию к этому явлению и кратко изложить. В чем по-вашему суть проклятия размерности и какое это имеет значение для задач машинного обучения?</p>
<p>Метрические методы машинного обучения</p>	<p>7. Метрические методы классификации. Виды расстояний</p> <p>8. Метод ближайшего соседа. Метод k ближайших соседей. Взвешенный метод k ближайших соседей</p> <p>9. Отбор эталонных объектов. Понятие отступа. Классификация объектов</p> <p>10. Регрессия по соседним объектам. Окно Парзена</p>	<p>1. Может ли в методе k ближайших соседей при $k = 2$ получиться лучший результат, чем при $k = 1$? Отказы от классификации тоже считать ошибками.</p>
<p>Линейные методы машинного обучения</p>	<p>11. Линейная регрессия. Постановка задачи. Матричная формулировка. Точное решение</p> <p>12. Регуляризация в линейной регрессии. Виды регуляризаторов</p> <p>13. Гребневая регрессия. Лассо-регрессия</p> <p>14. Итеративная версия решения задачи линейной регрессии без регуляризации и с регуляризацией. Метод градиентного спуска</p> <p>15. Вероятностная модель данных. Максимум правдоподобия</p> <p>16. Линейный классификатор. Отступы. Функционал качества</p> <p>17. Метод стохастического градиента (SGD). Преимущества и недостатки</p> <p>18. Методы улучшения сходимости SGD (метод моментов, метод Нестерова)</p> <p>19. Методы улучшения сходимости SGD (AdaGrad, RMSProp, Adam)</p> <p>20. Логистическая регрессия. Простой байесовский классификатор. Сигмоидальная функция</p> <p>21. Метод опорных векторов (SVM). Постановка задачи для линейно разделимой выборки без выбросов</p> <p>22. Отступы в SVM. Учет выбросов. Формулировка теоремы Куна-Таккера. Двойственная задача</p> <p>23. Классификация объектов в SVM. Постановка задачи через двойственные переменные λ</p> <p>24. Нелинейные ядра в методе SVM. Примеры ядер. Способы их построения</p> <p>25. Основные метрики качества алгоритмов. ROC-кривые</p>	<p>1. Покажите асимптотическую эквивалентность энтропийного и статистического критериев информативности.</p> <p>2. Какая стратегия поведения в листьях решающего дерева приводит к меньшей вероятности ошибки: отвечать тот класс, который преобладает в листе, или отвечать случайно с тем же распределением классов, что и в листе?</p> <p>3. Покажите, что регуляризатор в задаче линейной классификации имеет вероятностный смысл априорного распределения параметров моделей. Какие распределения задают l_1-регуляризатор и l_2-регуляризатор</p>
<p>Кластеризация и вероятностное моделирование данных</p>	<p>26. Методы кластеризации. Типы кластерных структур</p> <p>27. Функционал качества кластеризации.</p> <p>28. EM-алгоритм в кластерном анализе</p> <p>29. Метод k-средних</p> <p>30. Иерархическая кластеризация. Формула Ланса-Уильямса</p> <p>31. Быстрая агломеративная кластеризация.</p>	<p>Покажите, как получается условная оптимизационная задача, решаемая в SVM из соображений максимизации разделяющей полосы между классами. Можно отталкиваться от линейно разделимого случая, но итоговое выражение должно быть для общего. Как эта задача сводится к безусловной задаче оптимизации?</p>

Составитель (и): Решетникова Е.В., доцент