Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ ФГБОУ ВО «КемГУ»
Дата и время: 2024-02-21 00:00:00
471086fad29a3b30e244c728abc3661ab35c9d5021b0dcf0e75e03a5b6fdf6436

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
10 февраля 2022 г.

Рабочая программа дисциплины
К.М.07.ДВ.01.02 Глубокое 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 2022

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. 7
5.1 Учебная литература 7
5.2 Материально-техническое и программное обеспечение дисциплины. 7
5.3 Современные профессиональные базы данных и информационные справочные системы. 8
6 Иные сведения и (или) материалы. 9
6.1. Примерные вопросы и задания / задачи для промежуточной аттестации . 9

1 Цель дисциплины.

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

ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач

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
общепрофессиональная	Теоретические и практические основы профессиональной деятельности	*ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач*

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
ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК 2.1. Анализирует математические методы решения прикладных задач ОПК 2.2. Реализует математические методы решения прикладных задач ОПК 2.3 Модифицирует математические методы решения прикладных задач	К.М.02.04 Методы машинного обучения К.М.02.05 Методы математического прогнозирования К.М,02.06 Методы обработки многомерных данных К.М.06.02(Н) Научно-исследовательская практика К.М.07.01(Д) Выполнение и защита выпускной квалификационной работы

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
ОПК-2 Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК 2.1. Анализирует математические методы решения прикладных задач ОПК 2.2. Реализует математические методы решения прикладных задач ОПК 2.3 Модифицирует математические методы	**Знать*:*** – современные методы, используемые в машинном обучении – тенденции развития, научные и прикладные достижения в области машинного обучения **Уметь*:*** – - анализировать и модифицировать

Код и название компетенции	Индикаторы достижения компетенции, закрепленные за дисциплиной	Знания, умения, навыки (ЗУВ), формируемые дисциплиной
	решения прикладных задач	методы машинного обучения для определения наиболее подходящего при прогнозе конкретного процесса – использовать методы машинного обучения для решения научно-исследовательских и прикладных задач. **Владеть:** – 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	ОЗФО	ЗФО
1 Общая трудоемкость дисциплины	216		
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	64		
Аудиторная работа (всего):	64		
в том числе:			
лекции	32		
практические занятия, семинары			
практикумы			
лабораторные работы	32		
в интерактивной форме			
в электронной форме			
Внеаудиторная работа (всего):	116		
в том числе, индивидуальная работа обучающихся с преподавателем			
подготовка курсовой работы /контактная работа			
групповая, индивидуальная консультация и иные виды учебной деятельности, предусматривающие групповую или индивидуальную работу обучающихся с преподавателем)			
творческая работа (эссе)			
3 Самостоятельная работа обучающихся (всего)	116		
4 Промежуточная аттестация обучающегося и объём часов, выделенный на промежуточную аттестацию: - зачет 3 семестр, экзамен 4 семестр	36		

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 5 - Учебно-тематический план очной формы обучения

недели	Разделы и темы дисциплины	Общая трудоём	Трудоемкость занятий (час.)		Формы текущего
			ОФО	ОЗФО	

	по занятиям	кость *(всего час.)*	Аудиторн. занятия		СРС	Аудиторн. занятия		СРС	контроля и промежуточной аттестации успеваемости
			лекц.	лаб.		лекц.	практ.		
1-8	1. Введение в методы машинного обучения	36	8	8	20				Отчет по лабораторным работам
9-16	2. Метрические методы машинного обучения	36	8	8	20				Отчет по лабораторным работам
1-8	3. Линейные методы машинного обучения	54	8	8	38				Отчет по лабораторным работам
9-16	4. Кластеризация и вероятностное моделирование данных	54	8	8	38				Отчет по лабораторным работам
	Промежуточная аттестация (зачет 3 семестр, экзамен 4 семестр)	36							36
	Всего:	216	32	32	116				Зачет 3 семестр Экзамен 4 семестр

3.2. Содержание занятий по видам учебной работы

Таблица 6 – Содержание дисциплины

№ п/п	Наименование раздела, темы дисциплины	Содержание занятия
	Содержание лекционного курса	
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 обощающей 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 обощения. Возможные виды ядер
4	Кластеризация и вероятностное моделирование данных	Методы кластериации. Типы кластерных структур. Функционал качества кластеризации. EM-алгоритм. Метод k-средних. Иерархическая кластеризация. Формула Ланса-Уильямса. Быстрая агломеративная кластеризация.
	Содержание лабораторных занятий	
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)

3 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)	**5 баллов** посещение и конспект 1 лекционного занятия	15-40
		Лабораторные работы	**Отчет по лабораторной работе (4 отчета):** **5 баллов** (выполнено 51 - 65% заданий) **7 балла** (выполнено 66 - 85% заданий)4 **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.				

4 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)	**60**	Лекционные занятия (конспект) (8 занятий)	**2 баллов** посещение 1 лекционного занятия	8-16
		Лабораторные работы	**Отчет по лабораторной работе (4 отчета):** **5 баллов** (выполнено 51 - 65% заданий) **7 балла** (выполнено 66 - 85% заданий)4 **11 баллов** (выполнено 86 - 100% заданий)	18-44
Итого по текущей работе в семестре				41 - 60
Промежуточная аттестация (экзамен)	40	Теоретические вопросы (2 вопроса)	**10 баллов** (пороговое значение) **20 баллов** (максимальное значение)	10-20
		Решение задач(2 задачи).	**5 баллов** (пороговое значение) **10 баллов** (максимальное значение)	10-20
Итого по промежуточной аттестации (экзамен)				20-40
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 НФИ КемГУ:

410 Учебная аудитория (мультимедийная) для проведения: - занятий лекционного типа; - групповых и индивидуальных консультаций; - текущего контроля и промежуточной аттестации; Специализированная (учебная) мебель: доска меловая, кафедра, моноблоки аудиторные. Оборудование: стационарное - компьютер, экран, проектор. Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по сублицензионному договору № 1212/КМР от 12.12.2018 г. до 12.12.2021	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19

г.), LibreOffice (свободно распространяемое ПО), Яндекс.Браузер (отечественное свободно распространяемое ПО). Интернет с обеспечением доступа в ЭИОС.	
508 Лаборатория компьютерного моделирования Учебная аудитория (мультимедийная) для проведения: - занятий лабораторного типа; - самостоятельной работы; Специализированная (учебная) мебель: доска меловая, кафедра, столы, стулья. Оборудование для презентации учебного материала: стационарное - компьютер преподавателя, проектор, экран. Лабораторное оборудование: стационарное – компьютеры для обучающихся (18 шт.). Используемое программное обеспечение: MSWindows (MicrosoftImaginePremium 3 year по сублицензионному 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), Opera 12 (свободно распространяемое ПО), Python с расширениями PIL, Py OpenGL, (свободно распространяемое ПО) Интернет с обеспечением доступа в ЭИОС.	654079, Кемеровская область, г. Новокузнецк, пр-кт Металлургов, д. 19

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. Информационная система «Единое окно доступа к информационным ресурсам» http://window.edu.ru/catalog/
5. Базы данных и аналитические публикации на портале «Университетская информационная система Россия», режим доступа: https://uisrussia.msu.ru/
6. UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения - http://archive.ics.uci.edu/ml/
7. 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
8. Портал по интеллектуальному анализу данных, поддерживаемый Григорием Пятецким-Шапиро - http://www.kdnuggets.com/
9. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - http://machinelearning.ru

6 Иные сведения и (или) материалы.

6.1. 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	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) 5. Функционал качества. Минимизация эмпирического риска 6. Переобучение. Методы валидации моделей. Кросс-валидация	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). 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?
Метрические методы машинного обучения	7. Метрические методы классификации. Виды расстояний 8. Метод ближайшего соседа. Метод k ближайших соседей. Взвешенный метод k ближайших соседей 9. Отбор эталонных объектов. Понятие отступа. Классификация объектов 10. Регрессия по соседним объектам. Окно Парзена	1. Может ли в методе k ближайших соседей при $k = 2$ получиться лучший результат, чем при $k = 1$? Отказы от классификации тоже считать ошибками.
Линейные методы машинного обучения	11. Линейная регрессия. Постановка задачи. Матричная формулировка. Точное решение 12. Регуляризация в линейной регрессии. Виды регуляризаторов 13. Гребневая регрессия. Лассо-регрессия 14. Итеративная версия решения задачи линейной регрессии без регуляризации и с регуляризацией. Метод градиентного спуска 15. Вероятностная модель данных. Максимум правдоподобия 16. Линейный классификатор. Отступы. Функционал качества 17. Метод стохастического градиента (SGD). Преимущества и недостатки 18. Методы улучшения сходимости SGD (метод моментов, метод Нестерова) 19. Методы улучшения сходимости SGD (AdaGrad, RMSProp, Adam) 20. Логистическая регрессия. Простой байесовский классификатор. Сигмоидальная функция 21. Метод опорных векторов (SVM). Постановка задачи для линейно разделимой выборки без выбросов 22. Отступы в SVM. Учет выбросов. Формулировка теоремы Куна-Таккера. Двойственная задача	1. Покажите асимптотическую эквивалентность энтропийного и статистического критериев информативности. 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? 3. Покажите, что регуляризатор в задаче линейной классификации имеет вероятностный смысл априорного

	23. Классификация объектов в SVM. Постановка задачи через двойственные переменные λ 24. Нелинейные ядра в методе SVM. Примеры ядер. Способы их построения 25. Основные метрики качества алгоритмов. ROC-кривые	распределения параметров моделей. Какие распределения задают $l1$-регуляризатор и $l2$-регуляризатор
Кластеризация и вероятностное моделирование данных	26. Методы кластериации. Типы кластерных структур 27. Функционал качества кластеризации. 28. EM-алгоритм в кластерном анализе 29. Метод k-средних 30. Иерархическая кластеризация. Формула Ланса-Уильямса 31. Быстрая агломеративная кластеризация.	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?

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