

Подписано электронной подписью:
Вержицкий Данил Григорьевич
Должность: Директор КГПИ ФГБОУ ВО «КемГУ»
Дата и время: 2024-04-24 00:00:00
471086fad29a3b30e244e728abc3661ab35e9d50210dcf0e75e03a5b6fdf6436

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение высшего образования
«КЕМЕРОВСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Кузбасский гуманитарно-педагогический институт

Факультет информатики, математики и экономики

УТВЕРЖДАЮ
Декан
А.В. Фомина
«08» февраля 2024 г.

Рабочая программа дисциплины

К.М.09.06 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 2024

Новокузнецк 2024

Оглавление

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

1 Цель дисциплины.

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

Формируемые компетенции, индикаторы достижения компетенций, знания, умения, навыки

Таблица 1 – Индикаторы достижения компетенций, формируемые дисциплиной

Код и название компетенции	Индикаторы достижения компетенции по ОПОП	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	Знать: – современные методы, используемые в машинном обучении – тенденции развития, научные и прикладные достижения в области машинного обучения Уметь: – 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. Владеть: – навыками использования современных методов машинного обучения для решения прикладных задач

Место дисциплины

Дисциплина включена в модуль «Анализ данных» ОПОП ВО. Дисциплина осваивается на 4 курсе в 7 семестре.

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

Таблица 2 – 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	180
2 Контактная работа обучающихся с преподавателем (по видам учебных занятий) (всего)	37
Аудиторная работа (всего):	34
в том числе:	
лекции	6
лабораторные работы	28
Внеаудиторная работа (всего):	3
в том числе, индивидуальная работа обучающихся с преподавателем	
подготовка курсовой работы /контактная работа	3
3 Самостоятельная работа обучающихся (всего)	107
4 Промежуточная аттестация обучающегося и объём часов, выделенный на промежуточную аттестацию: - экзамен 7 семестр	36

3. Учебно-тематический план и содержание дисциплины.

3.1 Учебно-тематический план

Таблица 3 - Учебно-тематический план очной формы обучения

№ недели п/п	Разделы и темы дисциплины по занятиям	Общая трудоёмкость (всего час.)	Трудоёмкость занятий (час.)			Формы текущего контроля и промежуточной аттестации успеваемости
			ОФО		СРС	
			Аудиторн. занятия	лекц.		
1-8	1. Введение в методы машинного обучения	33	1	4	28	Отчет по лабораторным работам
9-16	2. Метрические методы машинного обучения	38	1	8	29	Отчет по лабораторным работам
1-8	3. Линейные методы машинного обучения	34	2	8	24	Отчет по лабораторным работам
9-16	4. Кластеризация и вероятностное моделирование данных	36	2	8	26	Отчет по лабораторным работам
	Промежуточная аттестация - курсовая работа, экзамен (7 семестр)	39				39
	Всего:	180	6	28	107	39

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 4.

Таблица 4 - Шкала и показатели оценивания результатов учебной работы обучающихся по видам в балльно-рейтинговой системе (БРС)

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 5)

Таблица 5 – Соотнесение 100-балльной шкалы и буквенного эквивалента оценки

Сумма набранных	Уровни освоения	Экзамен	Зачет
-----------------	-----------------	---------	-------

баллов	дисциплины и компетенций	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-Давид Ш. ; 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 (Microsoft Imagine Premium 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; 	<p>Учебный корпус №4.</p> <p>654079, Кемеровская</p>

<p>- самостоятельной работы; - текущего контроля и промежуточной аттестации.</p> <p>Специализированная (учебная) мебель: доска меловая, столы компьютерные, стулья.</p> <p>Оборудование для презентации учебного материала: <i>стационарное</i> - компьютер, экран, проектор, наушники.</p> <p>Лабораторное оборудование: стационарное – компьютеры для обучающихся (16 шт.).</p> <p>Используемое программное обеспечение: MSWindows (Microsoft Imagine Premium 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), Microsoft Visual Studio (Microsoft Imagine Premium 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>область, г. Новокузнецк, пр-кт Металлургов, 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 Иные 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 6 - 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) 5. Функционал качества. Минимизация эмпирического риска 6. Переобучение. Методы валидации моделей. Кросс-валидация 	<p>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?</p>
Метрические методы машинного обучения	<ol style="list-style-type: none"> 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 	<ol style="list-style-type: none"> 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	<ol style="list-style-type: none"> 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 23. Классификация объектов в SVM. Постановка задачи через двойственные переменные λ 	<ol style="list-style-type: none"> 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 распределения параметров моделей. Какие распределения задают $l1$-регуляризатор и $l2$-регуляризатор

	<p>24. Нелинейные ядра в методе SVM. Примеры ядер. Способы их построения</p> <p>25. Основные метрики качества алгоритмов. ROC-кривые</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