

МИНОБРНАУКИ РОССИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ
ВЫСШЕГО ОБРАЗОВАНИЯ
«ВОРОНЕЖСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
(ФГБОУ ВО «ВГУ»)

УТВЕРЖДАЮ

Заведующий кафедрой
Информационных технологий
и математических методов в экономике



И.Н. Щепина

18.04.2024 г.

РАБОЧАЯ ПРОГРАММА УЧЕБНОЙ ДИСЦИПЛИНЫ

Б1.В.ДВ.01.02 Прикладное машинное обучение на языке Python

- 1. Код и наименование направления подготовки/специальности:** 38.03.01 Экономика
- 2. Профиль подготовки/специализация:** Экономика, финансы, бизнес-аналитика
- 3. Квалификация выпускника:** бакалавр
- 4. Форма обучения:** очное
- 5. Кафедра, отвечающая за реализацию дисциплины:** Кафедра информационных технологий и математических методов в экономике
- 6. Составители программы:**
Каляпина Ольга Ивановна, к.т.н., доцент
- 7. Рекомендована:** НМС экономического факультета протокол №3 от 21.03.24 г.
- 8. Учебный год:** 2025/2026 **Семестр(ы):** 4

9. Цели и задачи учебной дисциплины

Целями освоения учебной дисциплины являются:

– теоретическая и практическая подготовка в области анализа данных с использованием языков программирования и специализированного ПО для моделирования бизнес-процессов, в том числе изучение принципов и методов поиска эффективных решений прикладных экономических задач по выбранным критериям;

Задачи учебной дисциплины:

– формирование навыков составления алгоритмов для описания экономических процессов;

– формирования навыков программирования для реализации составленных алгоритмов;

– изучение принципов и методов управления экономическими данными с применением информационных интеллектуальных технологий;

приобретение практических навыков работы в конкретной среде программирования.

10. Место учебной дисциплины в структуре ООП:

Дисциплина «Основы алгоритмизации и программирования» входит в часть учебного плана, формируемую участниками образовательных отношений блока Б1. Для ее освоения необходимы знания, умения и компетенции, сформированные в результате изучения дисциплин базовой части математического и естественнонаучного цикла. Дисциплина связана с дисциплинами: Математический анализ, Линейная алгебра, Теория вероятностей и математическая статистика, Информационные технологии в экономике.

11. Планируемые результаты обучения по дисциплине/модулю (знания, умения, навыки), соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями) и индикаторами их достижения:

Код	Название компетенции	Код(ы)	Индикатор(ы)	Планируемые результаты обучения
ПК-1	Способен описывать экономические процессы и системы, осуществлять сбор числовой и нечисловой информации, строить теоретические и экономико-математические модели, анализировать и обосновывать возможные решения	ПК-1.2	Собирает, систематизирует и анализирует данные в соответствии с требованиями заинтересованных сторон для формирования эффективных решений по выбранным критериям	<p>Знать:</p> <ul style="list-style-type: none">– принципы построения и верификации алгоритмов в соответствии с целями бизнес-анализа;– основные принципы сбора числовой и нечисловой информации в условиях цифровой экономики;– методы предварительной обработки данных (кодирование, стандартизация и нормализация, устранение выбросов, заполнение пропусков) <p>Уметь:</p> <ul style="list-style-type: none">– собирать числовую и нечисловую информацию с использованием современных информационных технологий;– анализировать многомерные данные для выявления тенденций и закономерностей;– преодолевать вычислительные проблемы, связанные с высокой размерностью данных;– составлять алгоритмы для поиска допустимых решений на основе собранных данных и целей бизнес-анализа <p>Владеть:</p> <ul style="list-style-type: none">– навыками построения и проверки качества алгоритмов функционирования экономических процессов;– навыками работы с цифровыми инструментами для сбора и обработки данных;

				– способностью к анализу и интерпретации результатов обработки информации для принятия обоснованных решений
ПК-2	Способен применять информационные технологии для проведения анализа социально-экономических процессов и систем, бизнес-анализа, подготовки информационно-аналитических материалов и прогнозирования экономических показателей	ПК-2.2	Готовит информационные обзоры, определяет основные тенденции развития, моделирует бизнес-процессы с использованием информационных технологий и оформляет результаты в виде аналитических отчетов	Знать: – один из основных языков программирования (Python, Java, C++, JavaScript и др.), используемых в экономическом анализе и моделировании, – специализированные программные продукты, такие как Excel, SPSS, R и другие. Уметь: – писать программы на выбранном языке программирования для обработки данных и создания моделей, – использовать специализированное программное обеспечение для проведения экономического анализа и моделирования. Владеть: – навыками разработки и применения программ на выбранном языке программирования, – навыками работы в специализированных программных продуктах для решения экономических задач и анализа данных.
		ПК-2.3	Использует языки программирования и специализированное ПО при решении прикладных экономических задач	Знать: – принципы и методы управления экономическими данными; – информационные интеллектуальные технологии для управления экономическими данными; – актуальные алгоритмы машинного обучения, используемые для решения экономических задач Уметь: – применять интеллектуальные информационные технологии для управления экономическими данными в соответствии с требованиями бизнеса. – использовать методы машинного обучения для автоматизации процессов управления экономическими данными, такими как сбор, обработка, анализ и хранение данных. 3. Владеть навыками: – анализа и обработки экономических данных с использованием информационных интеллектуальных технологий. – использования библиотек языка Python для автоматизации процессов анализа экономических данных

12. Объем дисциплины в зачетных единицах/час. (в соответствии с учебным планом) — 2/72.

Форма промежуточной аттестации (зачет/экзамен): зачет

13. Трудоемкость по видам учебной работы:

Вид учебной работы		Трудоемкость	
		Всего	4 семестр
Аудиторные занятия		32	32
в том числе:	лекции	-	-
	практические	-	-
	лабораторные	32	32
Самостоятельная работа		40	40
в том числе: курсовая работа (проект)		-	-
Форма промежуточной аттестации		зачет	зачет
Итого:		72	72

13.1. Содержание дисциплины:

п/п	Наименование раздела дисциплины	Содержание раздела дисциплины	Реализация раздела дисциплины с помощью онлайн-курса, ЭУМК*
Лабораторные занятия			
1	Обзор библиотеки Scikit-Learn языка программирования Python	Знакомство с библиотекой машинного обучения Scikit-Learn. Гиперпараметры и проверка качества модели.	-
2	Построение и отбор признаков модели	Извлечение признаков (Feature Extraction). Преобразование признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков. Выбор признаков (Feature selection): автоматические методы, визуализация, отбор с использованием моделей	https://edu.vsu.ru/course/view.php?id=24641#section-9
3	Решение задачи прогнозирования	Разбор примера построения модели прогнозирования для задачи предсказания трафика. Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества	-
4	Решение задачи прогнозирования для заданного набора данных	Построение модели прогнозирования с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели.	-
5	Алгоритмизация задачи классификации.	Разбор примера предсказания оттока клиентов мобильного оператора. Отбор и кодирование признаков. Проверка качества модели с помощью перекрёстной проверки.	-
6	Программирование задачи классификации	Разработка программы классификации с помощью библиотеки Scikit-Learn. Анализ качества построенной модели	-
7	Древовидные модели: деревья решений, случайный лес	Разбор примера построения модели дерева решений для задачи предсказания кредитоспособности клиента. Анализ деревьев, полученных при использовании различных метрик. Построение модели случайного леса на примере задачи кредитного скоринга. Кодирование признаков и заполнение пропущенных данных.	-
8	Программирование древовидной модели для	Построение моделей деревьев решений и случайного леса с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели	-

	заданного набора данных		
9	Программирование модели градиентного бустинга для заданного набора данных	Построение модели градиентного бустинга с помощью библиотеки Scikit-Learn для заданного набора данных. Анализ качества построенной модели	
10	Анализ текстовых данных	Разбор примера построения модели анализа текстовых данных для задачи определения тональности отзывов.	-

13.2. Темы (разделы) дисциплины и виды занятий

№ п/п	Наименование темы (раздела) дисциплины	Виды занятий (количество часов)				
		Лекции	Практические	Лабораторные	Самостоятельная работа	Всего
1	Введение в машинное обучение. Обзор экономических задач, решаемых с помощью ИИ	-	-	-	4	4
2	Алгоритмы прогнозирования	-	-	12	14	26
3	Алгоритмы классификации и кластеризации	-	-	12	14	26
4	Ансамблевые алгоритмы	-	-	4	4	8
5	Сверточные и рекуррентные нейронные сети. Генетические алгоритмы	-	-	4	4	8
	Зачет	-				
	Итого:	-	-	32	40	72

14. Методические указания для обучающихся по освоению дисциплины:

(рекомендации обучающимся по освоению дисциплины: указание наиболее сложных разделов, работа с конспектами лекций, презентационным материалом, рекомендации по выполнению курсовой работы, по организации самостоятельной работы по дисциплине и др.)

В процессе изучения дисциплины используются такие виды учебной работы, как лабораторные занятия, а также различные виды самостоятельной работы обучающихся.

Для подготовки к лабораторному занятию обучающийся должен заранее ознакомиться с заданием и теоретическим материалом, после выполнения работы оформить отчет о проделанной работе и подготовиться к ее защите. Все отчеты формируются в виде текстового файла и высылаются для проверки преподавателем.

При подготовке к лабораторным занятиям особое внимание следует уделять особенностям использования изучаемых программных продуктов и грамотному оформлению полученных результатов.

Самостоятельная работа обучающихся направлена на самостоятельное изучение отдельных тем и вопросов учебной дисциплины и является обязательной для каждого обучающегося, ее объем определяется учебным планом, обучающийся работает с рекомендованными материалами при минимальном участии преподавателя.

Самостоятельная работа с учебниками, учебными пособиями, научной, справочной и популярной литературой, материалами периодических изданий и ресурсами сети Internet, статистическими данными, написание программ по темам лабораторных работ является наиболее эффективным методом получения знаний, позволяет значительно активизировать процесс овладения информацией, способствует более глубокому усвоению изучаемого материала, формирует у обучающихся свое отношение к конкретной проблеме.

Вопросы, которые вызывают у обучающихся затруднения при подготовке, должны быть заранее сформулированы и озвучены во время занятий в аудитории для дополнительного разъяснения преподавателем.

Виды самостоятельной работы: конспектирование учебной и научной литературы; проработка учебного материала (по конспектам лекций, учебной и научной литературе); работа в электронной библиотечной системе; работа с информационными справочными системами, выполнение домашних заданий; выполнение контрольных заданий; подготовка к занятиям; работа с вопросами для самопроверки.

Рекомендованные методические материалы, задания к лабораторным работам, исходные данные для моделирования размещаются на странице курса «Основы алгоритмизации и программирования» на портале «Электронный университет ВГУ» <https://edu.vsu.ru/course/view.php?id=24641> автор Каляпина О.И.

15. Перечень основной и дополнительной литературы, ресурсов интернет, необходимых для освоения дисциплины (список литературы оформляется в соответствии с требованиями ГОСТ и используется общая сквозная нумерация для всех видов источников)

а) основная литература:

№ п/п	Источник
1.	Замятин И.В. Программирование на языке Python: учебно-методическое пособие / И.В.Замятин; Воронежский государственный университет.- Воронеж: Издательский дом ВГУ, 2019. – 33 с. То же [Электронный ресурс]. - URL: http://www.lib.vsu.ru/elib/texts/method/vsu/m19-160.pdf
2.	Хрипунова М.Б., Губернаторов А.М. Экономика на Python: учебник: - Москва : Прометей, 2022. – 316 с. - Режим доступа: https://e.lanbook.com/book/220832
3.	Рашка, С. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения [Электронный ресурс] : руководство / С. Рашка ; пер. с англ. Логунова А.В.. — Электрон. дан. — Москва : ДМК Пресс, 2017. — 418 с. — Режим доступа: https://e.lanbook.com/book/100905
4.	Козьло, Л.П. Построение систем машинного обучения на языке Python [Электронный ресурс] / Л.П. Козьло, В. Ричарт ; пер. с англ. Слинкин А. А.. — Электрон. дан. — Москва : ДМК Пресс, 2016. — 302 с. — Режим доступа: https://e.lanbook.com/book/82818
5.	Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных [Электронный ресурс] / П. Флах. — Электрон. дан. — Москва : ДМК Пресс, 2015. — 400 с. — Режим доступа: https://e.lanbook.com/book/69955

б) дополнительная литература:

№ п/п	Источник
6.	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с.
7.	Силен Дэви, Мейсман Арно, Али Мохамед. Основы Data Science и Big Data. Python и наука о данных. -СПб.: Питер, 2017. -336 с.
8.	Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с.:
9.	Шарден, Б. Крупномасштабное машинное обучение вместе с Python [Электронный ресурс] : учебное пособие / Б. Шарден, Л. Массарон, А. Боскетти ; пер. с англ. А. В. Логунова. — Электрон. дан. — Москва : ДМК Пресс, 2018. — 358 с. — Режим доступа: https://e.lanbook.com/book/10583
10.	Вьюгин, В.В. Математические основы машинного обучения и прогнозирования [Электронный ресурс] : учебное пособие / В.В. Вьюгин. — Электрон. дан. — Москва : МЦНМО, 2013. — 304 с. — Режим доступа: https://e.lanbook.com/book/56397
11.	Кук, Д. Машинное обучение с использованием библиотеки H2O [Электронный ресурс] / Д. Кук ; пер. с англ. Огурцова А.Б.. — Электрон. дан. — Москва : ДМК Пресс, 2018. — 250 с. — Режим доступа: https://e.lanbook.com/book/97353

в) информационные электронно-образовательные ресурсы (официальные ресурсы интернет)*:

№ п/п	Ресурс
12.	Академия Яндекса. Основы Python https://academy.yandex.ru/handbook/python
13.	Skillbox Media https://skillbox.ru/media/code/biblioteka-numpy-vsye-chto-nuzhno-znat-novichku/
14.	Бесплатные курсы по аналитике и Data Science https://netology.ru/free/data-science#/
15.	Онлайн платформа для проектов в области науки о данных https://www.kaggle.com/
16.	UCI Machine Learning Repository — репозиторий наборов данных для машинного обучения - http://archive.ics.uci.edu/ml/
17.	Ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. - http://machinelearning.ru
18.	Открытый курс машинного обучения https://habr.com/company/ods/blog/322626/

19.	Сайт официальной статистики Федеральной службы государственной статистики https://rosstat.gov.ru/folder/10705
-----	--

16. Перечень учебно-методического обеспечения для самостоятельной работы (учебно-методические рекомендации, пособия, задачки, методические указания по выполнению практических (контрольных), курсовых работ и др.)

№ п/п	Источник
1.	Курс «Основы алгоритмизации и программирования» на портале «Электронный университет ВГУ» https://edu.vsu.ru/course/view.php?id=24641 авторы Каляпина О.И., Щепина И.Н.
2.	Курс «Машинное обучение» на портале «Электронный университет ВГУ», автор Каширина И.Л. https://edu.vsu.ru/course/view.php?id=3579
3.	Бринк Х., Ричардс Д., Феверолф М. Машинное обучение. -СПб.: Питер, 2017. -336 с.: Материалы к книге: https://github.com/brinkar/real-world-machine-learning
4.	Плас Дж. Вандер Python для сложных задач: наука о данных и машинное обучение. — СПб.: Питер, 2018. — 576 с. Материалы к книге: https://github.com/jakevdp/PythonDataScienceHandbook
5.	А.Мюллер, С.Гвидо - Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными – 2017 электронный ресурс: https://owlweb.ru/vvedenie-v-mashinnoe-obuchenie-s-pomoshhyu-python-rukovodstvo-dlya-specialistov-po-rabote-s-dannymi/
6.	Хилпиш И. Python для финансовых расчетов, 2-е изд. : Пер. с англ. – СПб.: ООО «Диалектика», 2021. – 800 с. : ил. – Парал. тит. англ.

17. Образовательные технологии, используемые при реализации учебной дисциплины, включая дистанционные образовательные технологии (ДОТ), электронное обучение (ЭО), смешанное обучение): реализация учебной дисциплины предполагает применение дистанционных образовательных технологий (работу на образовательном портале «Электронный университет ВГУ»).

Python 3 с подключенными библиотеками (дистрибутив Anaconda)

18. Материально-техническое обеспечение дисциплины:

Лекционная аудитория должна быть оснащена современным компьютером с подключенным к нему проектором с видеотерминала на настенный экран. Лабораторные занятия должны проводиться в специализированной аудитории, оснащенной современными персональными компьютерами и программным обеспечением в соответствии с тематикой изучаемого материала.

19. Оценочные средства для проведения текущего контроля успеваемости и промежуточной аттестации

Порядок оценки освоения обучающимися учебного материала определяется содержанием следующих разделов дисциплины:

№ п/п	Наименование раздела дисциплины (модуля)	Компетенция(и)	Индикатор(ы) достижения компетенции	Оценочные средства
1.	Разделы 1-2	ПК-1, ПК-2	ПК–1.2, ПК–2.3	Задания для лабораторных работ 1-4
2.	Раздел 3	ПК-2	ПК–2.3	Задания для лабораторных работ 5-8
3.	Разделы 4-5	ПК-2	ПК–2.3	Задания для лабораторных работ 9-10
Промежуточная аттестация форма контроля – зачет				Индивидуальное задание к зачету

20 Типовые оценочные средства и методические материалы, определяющие процедуры оценивания

20.1 Текущий контроль успеваемости

Контроль успеваемости по дисциплине осуществляется с помощью следующих оценочных средств: выполнением заданий для лабораторных работ 1-10.

Текущие аттестации проводятся в соответствии с Положением о текущей аттестации обучающихся по программам высшего образования Воронежского государственного университета.

Лабораторная работа № 1

1. Скачайте с сайта Росстата <https://showdata.gks.ru/report/276220/> данные о заболеваемости населения (при желании можно использовать другие исходные данные).
2. Используя методы `read`, загрузите данные в датафрейм (при необходимости используйте параметр `encoding='cp1251'` для выбора кодировки данных).
3. Проанализируйте структуру датафрейма. При необходимости измените/удалите строки/столбцы/индексы.
4. Проанализируйте отсутствующие значения, обоснуйте и выберите метод заполнения отсутствующих значений. Заполните отсутствующие значения.
5. Проанализируйте данные в датафрейме – рассчитайте основные статистические показатели.
6. Используя сводные и перекрестные таблицы, проведите анализ показателей заболеваемости с разбивкой по регионам/годам.

Лабораторная работа № 2

1. Постройте графики для сводных показателей, вычисленных в ходе выполнения Лабораторной работы № 9. Выберите для этого наиболее подходящий вид графиков.
2. Сохраните построенные графики в виде графических файлов.

Лабораторная работа № 3

1. Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
2. Обучите, а затем провалидируйте на тестовых данных следующие модели, используя в качестве метрики качества R^2 , предварительно отмасштабировав данные
 - `LinearRegression`;
 - `Lasso` с коэффициентом регуляризации, равным 0.01.
3. Проанализируйте качество полученных моделей и сравните количество строго нулевых весов в них.

Лабораторная работа № 4

1. Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
2. Проведите предобработку данных: заполнение пропусков, кодирование, масштабирование
3. Обучите, а затем провалидируйте на тестовых данных модель логистической регрессии
4. Вычислите значения метрик: `recall`, `precision`, F1-мера, AUC-ROC. Постройте ROC-кривую. Сделайте выводы о качестве построенной модели.

Лабораторная работа № 5

Решите задачу бинарной классификации для предсказания уровня заработной платы по тексту объявления о вакансии на примере набора данных с соревнования на Kaggle. Данные доступны по ссылке.

1. Разбейте получившуюся выборку на обучающую и контрольную в соотношении 70/30
2. Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, осуществив векторизацию получившихся текстов при помощи `CountVectorizer`, обучив его на обучающей выборке и применив на тестовой.
3. Обучите логистическую регрессию из модуля `sklearn` с параметрами по умолчанию на обучающей выборке:
4. Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на тестовой выборке.
5. Отсортируйте веса признаков для модели. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.
6. Создайте текстовое описание объектов обучающей и контрольной выборок, объединив значения всех признаков каждого объекта выборки через символы пробела. После этого получите признаковое описание объектов, вычислив вектор `tf-idf` для каждого объекта помощи `TfidfVectorizer`, обучив его на обучающей выборке и применив на тестовой.
7. Заново обучите модель
8. Вычислите значения ROC-AUC, F-меры, а также постройте матрицу ошибок на контрольной выборке..

9. Сравните значения метрик из п. 8 со значениями, полученными в п. 4, и сравните соответствующие модели по качеству из работы.

10. Отсортируйте веса признаков для модели логистической регрессии из scikit-learn, полученной в п. 7. Какие слова из встречающихся в выборке имеют наибольшее/наименьшее влияние на значение целевой переменной? Проинтерпретируйте полученный результат.

20.2. Промежуточная аттестация

Для оценивания результатов обучения на зачете используются следующие показатели:

1) знание учебного материала и владение понятийным аппаратом описания экономических процессов посредством составления алгоритмов их функционирования;

2) умение анализа многомерные данные и преодолевать вычислительные проблемы, связанные с высокой размерностью данных;

3) умение применять методы машинного обучения при решении задач в различных прикладных областях;

5) владение навыками использования библиотек языка Python для построения систем, обучающихся по прецедентам

6) владение навыками построения и проверки качества моделей прикладных экономических задач;

7) владение навыками интерпретации полученных результатов в терминах экономического анализа и моделирования с целью получения новых знаний и выводов.

Для оценивания результатов обучения на зачете используется 2-балльная шкала: «зачтено», «незачтено».

Для сдачи **зачета** по дисциплине «Прикладное машинное обучение на языке Python» обучающиеся должны предоставить реферат на выбранную тему из следующего списка:

– Алгоритмы классификации. Метод опорных векторов (Support Vector Machines) и метод ближайших соседей.

– Алгоритмы предсказания.

– Ансамблевые алгоритмы. Бустинг, беггинг, стекинг.

– Алгоритмы кластеризации. Обучение без учителя: Mean-Shift, K-Means.

– Анализ экономической информации с помощью алгоритмов предсказания.

– Сверточные нейронные сети. Архитектура, параметры и алгоритмы.

– Рекуррентные нейронные сети. Алгоритмы LSM, LSTM, GRU.

– Обучение с подкреплением. Генетический алгоритм, Q-Learning, SARSA.

– Алгоритмы обобщения – сокращения размерности. Алгоритмы PCA, ICA

– Алгоритмы распознавания. Перцептрон. Метод обратного распространения ошибки.

Содержание реферата должно состоять из разделов:

– Введение (вопрос из списка ниже)

– Алгоритмы метода машинного обучения, архитектура сети

– Область применения

– Программная реализация одного из алгоритмов.

Вопросы для «Введения»:

– Что такое машинное обучение

– История машинного обучения

– Как связаны машинное и глубокое обучение, ИИ и нейросети

– Какие задачи решает машинное обучение

– Как устроено машинное обучение

– Основные виды машинного обучения

– Проблемы машинного обучения

– Перспективы машинного обучения: не начнет ли ИИ думать за нас?

– Типы задач машинного обучения

– Выбор параметров модели и кросс-валидация.

Объем реферата 10-15 стр, шрифт -Times New Roman 14, межстрочный интервал – одинарный.

20.3. Тестовые задания

Перечень заданий для проверки сформированности компетенций ПК-1, ПК-2:

1) тестовые задания (закрытого типа среднего уровня сложности):

Задание 1: Что подразумевается под интеллектуальным анализом данных?

- =извлечение ценной информации и знаний, которые могут быть неочевидными при традиционных методах анализа
- ~фильтрация данных с целью сокращения их размерности
- ~выявление скрытых закономерностей, трендов и связей в данных
- ~преобразование данных в другую форму представления

Задание 2: Основная характеристика задачи бинарной классификации:

- = зависимая переменная может принимать только два значения
- ~классификация осуществляется по двум признакам
- ~классификация осуществляется по одному признаку

Задание 3: Классификация относится к стратегии:

- =обучения без учителя
- ~обучения с учителем
- ~оба ответа неверны

Задание 4: Явление переобучения характеризуется ...

- =чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению
- ~ситуацией слишком долгого обучения из-за недостаточного числа обучающих примеров или слишком сложной структуры модели
- ~ситуацией слишком долгого обучения из-за простой структуры модели

Задание 5: Задача классификации сводится к ...

- =определению класса объекта по его характеристикам;
- ~нахождению зависимостей между объектами или событиями;
- ~определению по известным характеристикам объекта значения некоторого его параметра;
- ~поиску независимых групп и их характеристик во всем множестве анализируемых данных.

Задание 6: Множество примеров, используемое для конструирования модели, называется...

- =обучающим множеством
- ~тестовым множеством
- ~валидационным множеством

Задание 7: Преобразование всех символов строки в верхний регистр осуществляется по команде

- =`uppercase_string = "hello world".upper()`
- ~ `convert_to_lower = lambda text: text.lower()`
- ~`sum_of_elements = sum([1, 2, 3, 4, 5])`
- ~`is_element_in_list = "a" in ["a", "b", "c"]`
- ~ `number_to_string = str(5)`

Задание 8: Какое действие выполняет команда `index_of_element = "c".index("c")`?

- = Определение индекса элемента в списке
- ~ Создание нового списка на основе старого с измененным элементом
- ~ Удаление элемента из списка
- ~ Удаление символов из строки