

ПРОГРАММА ВСТУПИТЕЛЬНОГО ИСПЫТАНИЯ

по предмету «Технологии искусственного интеллекта и Big Data»
для поступающих на основную образовательную программу магистратуры
ВМ.5893 «Технологии искусственного интеллекта и Big Data»
по направлению подготовки
02.04.02 «Фундаментальная информатика и информационные технологии»

Раздел 1. Содержание теоретической части экзамена

Тема 1. Основы машинного обучения

1. Постановка задачи классификации. Примеры задач классификации. Типы алгоритмов классификации.
2. Метрические алгоритмы классификации. k-NN.
3. Логистические методы классификации. Деревья решений.
4. Постановка задачи кластеризации. Примеры задач кластеризации. Типы алгоритмов кластеризации.
5. Алгоритм Иерархической кластеризации. Алгоритм k-средних.

Тема 2. Теория вероятностей и математическая статистика

1. Основные понятия теории вероятностей и математической статистики (отношения между событиями. Невозможное, достоверное, противоположное события. Независимость случайных событий. Попарная независимость и независимость в совокупности).
2. Понятие случайной величины. Классификация случайных величин.
3. Функция распределения случайной величины, ее свойства.
4. Числовые характеристики случайных величин. Начальный и центральный моменты. Математическое ожидание, его свойства. Дисперсия случайной величины, ее свойства.
5. Элементы корреляционного анализа. Парная линейная регрессия.

Тема 3. Основы программирования на Python

1. Основные синтаксические конструкции, REPL.
2. Jupyter. Организация работы. Типовая организация проекта, виртуальные окружения, управление зависимостями pip, pip-env, poetry.
3. Аннотации типов, статистический анализ.
4. Проблемы интерпретатора, управление памятью, GIL.
5. Многопроцессорная обработка данных (Multiprocessing).

Тема 4. Базы данных

1. Основные понятия реляционной и объектной моделей данных.
2. Организация и проектирование физического уровня БД. Ключи. Методы индексирования.
3. Проектирование реляционных баз данных с использованием семантических моделей.
4. Обобщенная архитектура, состав и функции системы управления базой данных (СУБД). Характеристика современных технологий БД. Примеры соответствующих СУБД.
5. Язык SQL. Средства определения и изменения схемы БД, определения ограничений целостности. Контроль доступа. Средства манипулирования данными.

Тема 5. Методологические вопросы разработки ПО

1. Определение методологии разработки ПО, ее свойства и запись. Водопадная модель разработки. Agile Manifesto. Принципы agile. Гибкие методологии разработки ПО.
2. Процесс разработки ПО и организация работы команды. Водопадная модель, V-модель, инкрементальная, спиральная модели.
3. Гибкие методологии. Agile-манифест. История и основные положения экстремального программирования. Современные гибкие методологии.

Тема 6. Инструменты разработки ПО

1. Командные оболочки, языки сценариев командных оболочек. Командные оболочки различных семейство операционных систем (Windows, Unix-подобных).
2. Системы программирования (СП), типовые компоненты СП: языки, трансляторы, редакторы связей, отладчики, текстовые редакторы. Модульное программирование. Типы модулей. Связывание модулей по управлению и данным.
3. Системы управления версиями. Централизованное и распределённое управление версиями. Git Workflow, GitHub Workflow.

Раздел 2. Примеры практических заданий

1. Дана схема реляционной базы данных, например авиаперевозок пассажиров, в которой содержится информация о пассажирах и стоимости проданных билетов. Необходимо построить SQL-запрос, который позволяет вычислить распределение сумм забронированных билетов, входящих в заданные диапазоны стоимости билетов (с шагом 100 тысяч рублей).
2. Дан файл в заданном формате (например, в формате CSV). Требуется с помощью языка программирования Python написать программу, которая извлечет из данного файла всю текстовую информацию и вычислит сумму частот (классический term frequency) пяти самых популярных слов в тексте. Примечание: при вычислении суммы частот слов необходимо все слова в тексте приводить к единой нормальной форме.

Раздел 3. Информационные источники

Список рекомендованной литературы

1. An Introduction to Statistical Learning with Applications in R / G. James, D. Witten. - Corrected at 4 printing 2014. - New York : Springer, 2013.
2. Data mining: practical machine learning tools and techniques / I. H. Witten, E. Frank, M. A. Hall. - 3th. Ed. – eBook., База данных: eBook Collection (EBSCOhost).
3. Основы data science и big data. Python и наука о данных: учебное пособие / Д. Силен, А. Мейсман, М. Али. - СПб.: Питер, 2017.
4. The Elements of Statistical Learning: Data Mining, Inference, and Prediction. Trevor Hastie, Robert Tibshirani, Jerome Friedman. Series: Springer Series in Statistics. Edition: Second edition, corrected 7th printing. New York : Springer. 2009.
5. Марк Лутц. Изучаем Python Т.1// М.: ИД «Вильямс». – 2019. – С. 832.
6. Марк Лутц. Изучаем Python Т.2// М.: ИД «Вильямс». – 2020. – С. 720.
7. Разработка программного обеспечения: учебное пособие / Л. Константайн, Л. Локвуд ; пер. с англ.: В. Шрага. - СПб.: Питер, 2004. - 592 с.
8. Технология разработки программного обеспечения [Текст]: учебное пособие / Э. Дж. Брауде; ред., рук. работы Е. Строганова, ред. Ф. А. Новиков. - СПб. : Питер, 2004. - 654 с.

9. Буре В. М., Парилина Е. М. Теория вероятностей и математическая статистика: Учебник.– Лань, 2013.– 416 с.
10. Емельянов Г. В., Скитович В. П. Задачник по теории вероятностей и математической статистике. — 2-е изд., стер. — Спб.: Лань, 2007. — 336 с.
11. Зубков А.М., Севастьянов Б.А., Чистяков В.П. Сборник задач по теории вероятностей — М.: Наука, 1989.
12. Т.Коннолли, К.Бегг Базы данных. Проектирование, реализация и сопровождение. Теория и практика. (3-е издание). - М.:Вильямс, 2003. - 1440 с.
13. Введение в системы баз данных: учебное пособие / К. Дж. Дейт; пер. с англ. и ред.: К. А. Птицын. – 8-е изд. – М.; СПб.; Киев: Издательский дом "Вильямс", 2005. – 1328 с.
14. Системы баз данных. Полный курс / Гектор Гарсиа-Молина, Джефффри Д. Ульман, Дженнифер Уидом – М.: Вильямс, 2014. – 1088 с.
15. Сомервилл И. Инженерия программного обеспечения. М.: Вильямс, 2002.
16. Липаев В.В. Программная инженерия. Методологические основы. М.: Государственный Университет – Высшая школа экономики, 2006.

Список дополнительной литературы

1. Профессиональный информационно-аналитический ресурс, посвященный машинному обучению, распознаванию образов и интеллектуальному анализу данных. MachineLearning.ru – www.machinelearning.ru/wiki/index.php?title=Заглавная_страница
2. Буре В. М., Парилина Е. М., Седаков А. А. Методы прикладной статистики в R и Excel. / учеб. пособие.– Лань, 2016.– 152 с.
3. Боровков А. А. Математическая статистика. — М.: Изд. Наука, 1984.
4. Севастьянов Б. А. Курс теории вероятностей и математической статистики. — М.: Изд. Наука, 1982.
5. Феллер В. Введение в теорию вероятностей и ее приложения. Т. 1, 2. — М.: Изд. Мир, 1984.
6. The Art of Scrum How Scrum Masters Bind Dev Teams and Unleash Agility / McKenna, Dave; Berkeley, CA : Apress : Imprint: Apress, 2016.
7. Когаловский М.Р. Энциклопедия технологий баз данных. М.: Финансы и статистика, 2002.

Перечень иных информационных источников

1. Прикладные задачи анализа данных, Дьяконов <https://github.com/Dyakonov/PZAD/blob/master/README.md>
2. Документация Python 3.10.0: <https://docs.python.org/3/>
3. Pro Git book, интернет-ресурс. <https://git-scm.com/book/ru/v2>
4. The Scrum Guide, интернет-ресурс. <https://www.scrum.org/resources/scrum-guide>

Раздел 4. Организационно-методический раздел

Длительность вступительного испытания

180 минут (3 астрономических часа).

Форма вступительного испытания

Вступительный экзамен проводится в письменной форме без использование каких-либо информационных источников во время экзамена.

В случае выполнения в электронной форме с использованием программных средств, не предоставляющих возможности ввода формул и форматированного текста, математические формулы рекомендуется вводить в формате (La)TeX.

Структура и содержание вступительного испытания

Экзаменационное задание включает три раздела.

- 1. Научно-исследовательская часть.** Абитуриент представляет в обезличенной (не допускающей установление его/её личности) форме, во время экзамена, в поле, отведенном для ответа на вопрос в текстовом виде:
 - а. опишите текущий научный и профессиональный опыт,
 - б. аргументируйте выбор данной образовательной программы,
 - в. опишите возможности применения полученных знаний в будущей профессиональной деятельности;
 - г. напишите аннотацию своей выпускной работы, полученной на предыдущем уровне образования, с раскрытием её актуальности и перечислением основных результатов работы.
- 2. Теоретическая часть.** Абитуриент письменно отвечает на два произвольно выбранных экзаменационной комиссией вопроса по разным темам (см. Раздел 1).
- 3. Практическая часть.** Абитуриент получает одну практическую задачу, аналогичную приведённым примерам, и решает её с использованием языка программирования Python или SQL (см. Раздел 2).

Рекомендуемый объём: по 1 части — 300–600 слов, по 2 части (в совокупности по двум вопросам) — 400–800 слов.

Критерии оценивания вступительного испытания

Подсчет итоговой оценки за вступительное испытание осуществляется путем суммирования баллов, выставленных за ответы по каждому разделу. Вступительное испытание оценивается по шкале от 0 до 100 баллов.

Научно-исследовательская часть

Критерий	Максимальный балл
Описание текущего профессионального и/или научного опыта, релевантного направлению подготовки по данной образовательной программе	3
Наличие аргументов выбора данной образовательной программы	2
Соответствие приведенных возможностей полученных знаний профессиональной деятельности соответствующей образовательной программе	2
Ясность изложения аннотации ВКР, актуальность темы	7
Согласованность основных результатов ВКР	6
Соответствие тематики и результатов ВКР тематике выбранной образовательной программы	10

Итого: от 0 до 30 баллов по данной части.

Теоретическая часть

Ответы на два вопроса оцениваются отдельно, от 0 до 20 баллов каждый.

Каждый из двух вопросов оценивается следующим образом:

Качество ответа	Количество баллов
Полный корректный ответ	20
Незначительные локальные неточности, опечатки/описки	15
Неточности, не нарушающие ход рассуждения и изложения	10
Неточности, влияющие на ход рассуждения или изложения	5
Ответ, демонстрирующий непонимание смысла вопроса, его тематики	0
Ответ с содержательными (в т.ч. математическими) ошибками	0

Итого: от 0 до 40 баллов по данной части.

Практическая часть

Критерий	Максимальный балл
Корректность понимания задания, выбор подходящего алгоритма	5
Работоспособность программного кода, в т.ч. на корректных, но вырожденных (например, пустых) входных данных	15
Аккуратность оформления программного кода	10

Итого: от 0 до 30 баллов по данной части.