

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Романчук Иван Сергеевич
Должность: Ректор
Дата подписания: 18.11.2024 18:57:09
Уникальный программный ключ:
6319edc2b582ffdacea443f01d5779368d0957ac34f5cd074d81181530452479

МЕТОДИЧЕСКИЕ РЕКОМЕНДАЦИИ ПО ОРГАНИЗАЦИИ САМОСТОЯТЕЛЬНОЙ РАБОТЫ ОБУЧАЮЩИХСЯ

Наименование дисциплины	Анализ данных в гуманитарных науках / Data Analysis in Humanities
Направление подготовки	<i>45.04.02 Лингвистика</i>
Направленность (профиль)	<i>Прикладная лингвистика</i>
Форма обучения	<i>очная</i>
Разработчик	<i>Ганжерли Н.В., старший преподаватель</i>

1. Темы дисциплины для самостоятельного освоения обучающимися
Отсутствуют.

2. План самостоятельной работы

№ п/п	Учебные встречи	Виды самостоятельной работы	Форма отчетности/ контроля	Количество баллов	Рекомендуемый бюджет времени на выполнение (ак.ч.)*
1	2	3	4	5	6
1	УВ 2. «Язык программирования Python»	Составление терминологического глоссария по операциям, типам данных, правилам оформления кода.	Термины. Карточки с примерами кода с изученными операциями, типами данных, оформленные по правилам PEP8.	3	6
2	УВ 4. «Нормализация текстовых данных»	Составление глоссария, составление собственного протокола нормализации текстовых данных.	Термины.	1	6
3	УВ 9. «Частотное распределение»	Составление глоссария. Организация понятий глоссария в понятийную сетку.	Термины.	1	4
4	УВ 11, 17. «Отклонение от среднего значения», «Нормальное распределение»	Составление глоссария по новым библиотекам, использованным на практических занятиях.	Термины.	1	4
5	УВ 12. «Визуализация особенностей распределения»	Составление глоссария по новым библиотекам, использованным на практических занятиях.	Термины.	1	4
6	УВ 13. «Базовые понятия теории вероятности»	Составление глоссария и понятийной сетки.	Термины.	1	4
7	УВ 15. «Линейная регрессия»	Составление глоссария и понятийной сетки.	Термины.	1	4
8	УВ 16. «Наивный классификатор Байеса»	Составление глоссария и понятийной сетки.	Термины.	1	5
9	УВ 20. «Г-критерий	Составление	Термины.	1	4

	Стьюдента»	гlossария и понятийной сетки.			
10	УВ 21. «Хи- квадрат»	Составление гlossария и понятийной сетки.	Термины.	1	5

3. Требования и рекомендации по выполнению самостоятельных работ обучающихся, критерии оценивания

Составление гlossария и понятийной сетки.

- 1) Определите тему
- 2) Обратитесь к справочной литературе (не менее двух источников) и выделите ключевые термины (5-10).
- 3) Занесите определение термина в гlossарий.
- 4) Определите связи между терминами (родовидовые, причинно-следственные, иерархии, ассоциации или другие типы связей).
- 5) (по возможности: Создайте визуальную структуру. Используйте диаграммы или графические программы (XMind, MindMeister, MindMap, PowerPoint), чтобы визуализировать связи между терминами. Начните с центральной концепции и добавляйте ветви для каждого связанного термина, обозначая связи стрелками или линиями. При отсутствии технической возможности оформите термины на небольших карточках (лицевая сторона – термин, оборот – определение), сгруппируйте их как указано выше в этом пункте.).
- 6) Добавьте определения и примеры для каждого термина. Это поможет лучше понять и запомнить материал.
- 7) Организуйте понятийную сетку логически. Убедитесь, что информация организована логично и последовательно: постарайтесь рассказать изученный материал с опорой на полученную понятийную сетку.
- 8) Определите мета-связи между разными темами, чтобы можно было проследить их взаимосвязь.

Для оценивания СРС рекомендуется предоставить список терминов по теме и полученные визуализации.

4. Рекомендации по самоподготовке к промежуточной аттестации по дисциплине

При проведении промежуточной аттестации результаты, полученные обучающимся в семестре, переводятся в формат традиционной оценки в соответствии со шкалой перевода баллов:

- 60 баллов и менее – «неудовлетворительно»;
- от 61 до 75 баллов – «удовлетворительно»;
- от 76 до 90 баллов – «хорошо»;
- от 91 до 100 баллов – «отлично».

Если по результатам работы в семестре студент хочет улучшить оценку, он также может сдать экзамен. При этом баллы за семестр аннулируются. Когда студент выбирает вариант сдачи экзамена, работа за семестр принимается во внимание, но не учитывается при выставлении финальной оценки, которая может быть как выше оценки, полученной по итогам работы за семестр, так и ниже.

Экзамен проводится в форме письменного изложения знаний, которые есть у студента по одной из тем (список тем приводится далее), и имеет целью выявление уровня освоения дисциплины, характеризующего знания обучающегося в соответствии с определенными компетенциями.

Темы для подготовки к промежуточной аттестации:

1. Данные: количественные: дискретные, непрерывные, порядковые; качественные: номинальные, порядковые.

2. Обработка лингвистических данных при помощи компьютерных программ: нормализация, лемматизация, стемминг для небольшого корпуса текстов на русском и английском языках.

3. Распределение языковых единиц в корпусах текстов. Распределение слов, стемов и словоформ. Закон Ципфа. Туре-token ratio. Униграммы, биграммы, гапаксы.

4. Параметризация языковых единиц. Генеральная совокупность и выборка.

5. Переменные. Виды переменных. Шкала переменных. Группирующие переменные. Ранжирование. Статистическая значимость.

6. Показатели центральной тенденции: среднее арифметическое, среднее гармоническое, мода, медиана.

7. Случайная величина: дискретная, непрерывная. Вероятность абсолютная, относительная. Сумма, умножение вероятностей.

8. Мера рассеяния признака. Показатели меры рассеяния признака: лимиты, вариационный размах, среднее линейное отклонение.

9. Дисперсия, среднее квадратичное отклонение, коэффициент вариации, квадратичная ошибка средней.

10. Корреляция. Коэффициент корреляций. Коэффициент корреляции Пирсона.

11. Симметрия, эксцесс. Критерий Шапиро-Уилка. Критерий хи-квадрат.

12. Функция распределения непрерывной случайной величины.

13. Функция плотности распределения вероятностей. Плотность вероятности.

14. Нормальный закон распределения вероятностей. Нормальное распределение.

15. Проверка статистических гипотез.

16. Ошибки первого и второго рода. Уровень значимости. Статистическая мощность.

Квантиль.

17. Распределение хи-квадрат. Критерий согласия Пирсона.

18. Т-критерий Стьюдента.

19. t-статистика. Нулевая гипотеза.

20. Оценка согласованности экспертов.

21. Альфа Криппендорфа. Каппа Фляйса.

Требования к письменному ответу:

- Полнота освещения темы (Логическая структура: наличие введения, основной части и заключения. Ученые, внесшие вклад в создание и развитие области.).

- Ясность и четкость формулировок. Полное изложение теоретических основ.

- Связность и организация идей, представленных в логической последовательности. Исчерпывающий терминологический аппарат с необходимыми определениями.

- Примеры, адекватное оперирование терминологическим аппаратом; обсуждение стандартных и пограничных случаев.

Высшим баллом оценивается работа, в полной мере отвечающая указанным требованиям. При наличии терминологических неточностей оценка может быть снижена. Грубая ошибка (в терминах, применимости) значительно снижает общий балл. Наличие нескольких (двух и более) грубых ошибок позволяет выставить оценку «неудовлетворительно».

Список источников:

- Головин, Борис Николаевич. Язык и статистика [Текст] / Б. Н. Головин. – Москва: Просвещение, 1971. – 189 с.

- Маккинли, Уэс. Python и анализ данных / Уэс Маккинли ; перевод А. Слинкина. – 2-е изд. – Саратов: Профобразование, 2019. — 482 с. — ISBN 978-5-4488-0046-7. — Текст:

электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. — URL: <https://www.iprbookshop.ru/88752.html> (дата обращения: 06.05.2024). — Режим доступа: для авторизир. пользователей

- Грас, Джоэл. Data Science. Наука о данных с нуля [Текст] / Джоэл Грас; [пер. с англ. Андрея Логунова]. — 2-е изд., перераб. и доп. — Санкт-Петербург: БХВ-Петербург, 2021. - 416 с.: ил.; 24 см.; ISBN 978-5-9775-6731-2.

- Joel Grus. Data Science from Scratch. 2nd edition. O'Reilly Media, Inc. — 2019. ISBN 9781492041139.

- <https://www.statisticshowto.com/>