Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Новосибирский государственный технический университет»
Кафедра вычислительной техники

Паспорт контрольной работы

по дисциплине «Анализ и интерпретация данных», 3 семестр

1. Методика оценки

Выполнение контрольной работы является формой текущей аттестации (контроля) по дисциплине, предусмотренной учебным планом.

Контрольная работа направлена на закрепление и проверку уровня владения учебным материалом по теоретическим темам и темам практических занятий, а также на формирование навыков самостоятельного выбора методов и проведения анализа и интерпретации данных.

Номер варианта назначается преподавателем из предложенного списка вариантов. Также возможен выбор темы контрольной работы, связанный с тематикой магистерской диссертации (по согласованию с преподавателем).

Структура контрольной работы:

1. Введение
2. Обзорная часть. Описание выбранного метода анализа данных
3. Основная часть
3.1. Постановка задачи
3.2. Описание хода и результатов статистического исследования данных
4. Выводы, интерпретация полученных результатов

Список литературы

Приложение (в случае необходимости)

Во введении формулируется и кратко обосновывается актуальность, цель и задачи исследования, дается краткая характеристика степени изученности вопроса.

Обзорная часть – это описание метода решения задачи анализа данных.

Основная часть – это описание статистического исследования заданных выборочных данных с обоснованием выбора методов и программных средств, анализом и интерпретацией полученных результатов. Основная часть должна содержать обязательные ссылки на используемую литературу, интернет-источники, оформленные в соответствии с библиографическими требованиями.

В качестве исходных данных для контрольной работы используются многомерные выборки реальных данных из разных предметных областей. Приветствуется, если данные для статистической обработки предлагает студент.

Заключение: изложение общих выводов по результатам проведенного исследования, интерпретация полученных результатов в терминах предметной области.

Список литературы оформляется в соответствии с библиографическими требованиями в алфавитном порядке и включает от 3 до 5 источников (книг, статей разных авторов, интернет-источников, документов), которые были изучены при выполнении контрольной работы.

Рекомендуется излагать мысли по существу, кратко и логично.

Требования к оформлению:

Объем контрольной работы до 20 страниц машинописного текста формата А4. Шрифт Times New Roman, 12. Нумерация страниц сквозная, в нижней части листа по центру арабскими цифрами. Контрольная работа должна быть отредактирована, не содержать орфографических, синтаксических и стилевых ошибок.

Контрольная работа предоставляется для проверки в электронном виде в срок, установленный преподавателем. При положительном результате студент её защищает до сессии в назначенное преподавателем время. По результатам защиты студенту выставляется оценка в соответствии с критериями, приведенными в п. 2 настоящего Паспорта.

В контрольной работе проводится статистическое исследование выборочных данных, в ходе которого необходимо:

1. Составить план проведения статистического исследования, определить его цели, задачи, исходные данные, последовательность этапов, результаты, которые должны быть получены.
2. Решить типовую задачу АД (в зависимости от варианта задача кластеризации, классификации или прогнозирования) в среде выбранного статистического пакета или с использованием выбранного программного обеспечения.
3. Сделать выводы, интерпретировать полученные результаты в терминах прикладной задачи.
4. По результатам статистического исследования составить научно-исследовательский отчет о проделанной работе в соответствии с требованиями ГОСТ 7.32-2001.

2. Критерии оценки результатов обучения, соотнесенных с уровнями освоения индикаторов достижения компетенций

Общие правила выставления оценок текущей аттестации определяются балльно-рейтинговой системой, установленной локальным актом НГТУ.

Контрольная работа выполнена **на продвинутом** уровне, если структура, содержание и оформление работы соответствует требованиям. Все части контрольной работы согласованы, текст логично выстроен и является авторским. Присутствуют ссылки на нормативные документы и актуальную литературу. Работа представлена для проверки в установленные сроки. Анализ каждого из разделов контрольной работы свидетельствует о том, что совокупность результатов ее выполнения и соотнесенных с ними индикаторов достижения компетенций не содержит пробелов. Закрепленные за контрольной работой компетенции сформированы на продвинутом уровне. Оценка составляет *от 18 до 20 баллов*.

Контрольная работа выполнена **на базовом** уровне, если структура, содержание и оформление работы соответствует требованиям, но работа содержит единичные не принципиальные ошибки, исправленные после замечаний преподавателя. Все части контрольной работы согласованы, текст логично выстроен и является авторским. Присутствуют ссылки на нормативные документы и актуальную литературу. Работа представлена для проверки в установленные сроки. Анализ каждого из разделов контрольной работы свидетельствует о том, что совокупность результатов ее выполнения и соотнесенных с ними индикаторов достижения компетенций содержит несущественные пробелы. Закрепленные за контрольной работой компетенции сформированы на базовом уровне. Оценка составляет *от 15 до 17 баллов*.

Контрольная работа выполнена **на пороговом** уровне, если структура, содержание и оформление работы соответствует требованиям, но работа содержит ошибки, неоднократно исправляемые после замечаний преподавателя. Части контрольной работы в

целом согласованы. Присутствуют ссылки на нормативные документы и актуальную литературу. Работа представлена для проверки в установленные сроки. Анализ каждого из разделов контрольной работы свидетельствует о том, что совокупность результатов ее выполнения и соотнесенных с ними индикаторов достижения компетенций содержит пробелы. Закрепленные за контрольной работой компетенции сформированы на пороговом уровне. Оценка составляет *от 10 до 14 баллов*.

Контрольная работа считается **не выполненной,** если структура, содержание и оформление работы не соответствует требованиям, работа содержит существенные ошибки, не исправленные после замечаний преподавателя. Части контрольной работы не согласованы. Отсутствуют ссылки на нормативные документы и актуальную литературу. Работа не представлена для проверки в установленные сроки. Анализ каждого из разделов контрольной работы свидетельствует о том, что совокупность результатов ее выполнения и соотнесенных с ними индикаторов достижения компетенций не содержит множество существенных пробелов. Закрепленные за контрольной работой компетенции не сформированы. Оценка составляет *от 0 до 9 баллов*.

3. Шкала оценки

Контрольная работа как форма текущей аттестации (контроля) по дисциплине считается успешно выполненной, если сумма полученных баллов по всем ее заданиям составляет от 10 до 20 баллов включительно.

В общей оценке по дисциплине баллы за выполнение контрольной работы учитываются в соответствии с правилами балльно-рейтинговой системы и таблицей соответствия баллов, традиционной оценки и буквенной оценки ECTS, установленными в НГТУ.

4. Примерный перечень заданий (вариантов) контрольной работы

Варианты:

1. Кусочно-полиномиальная регрессия
2. Гребневая регрессия, метод лассо
3. Сглаживающие сплайны
4. Анализ главных компонент
5. Факторный анализ
6. Анализ пропущенных значений
7. Построение ансамбля моделей классификации (случайные леса)
8. Методы бустинга регрессионных деревьев
9. Методы бэггинга регрессионных деревьев
10. Построение ансамбля моделей кластеризации
11. Построение ансамбля моделей прогнозирования временных рядов
12. Самоорганизующиеся карты Кохонена
13. Методы сравнения точности моделей
14. Нейросетевые методы для решения задач прогнозирования
15. Нейросетевые методы для решения задач классификации
16. Нейросетевые методы для решения задач кластеризации
17. Обработка пропущенных наблюдений во временных рядах
18. Выявление структурных изменений во временных рядах
19. Байесовские структурные модели временных рядов
20. Кластеризация временных рядов
21. Анализ последовательностей событий

22. Метод опорных векторов
23. Анализ сетей (модели случайных сетей)
24. Анализ сетей (статистические модели сетей)
25. Анализ сетей (динамические модели сетей)
26. Методы обезличивания данных
27. Методы обогащения данных
28. Исследование сетевого трафика
29. Исследование методов предобработки данных для улучшения точности моделей машинного обучения.
30. Применение неоднородных ансамблей для повышения точности решения задачи классификации (стекинг моделей).
31. Применение неоднородных ансамблей для повышения точности решения задачи регрессии (стекинг моделей).
32. Применение неоднородных ансамблей для повышения точности решения задачи кластеризации (стекинг моделей).
33. Исследование методов обработки несбалансированных классов в задачах классификации.
34. Оптимизация гиперпараметров модели машинного обучения с использованием алгоритмов оптимизации.
35. Анализ и сравнение алгоритмов детекции аномалий для выявления нестандартных паттернов в данных.
36. Исследование методов сжатия данных для уменьшения размера наборов данных при сохранении важной информации (методы понижения размерности, компрессия данных, семплирование, агрегация, …).
37. Оценка и сравнение методов декомпозиции данных с целью выделения факторов и компонент (PCA, факторный анализ, ICA, спектральное разложение, …).
38. Применение ансамблевых методов (бэггинг и/или бустинг, необходимо рассмотреть минимум 2 алгоритма) для улучшения качества классификации.
39. Применение ансамблевых методов (бэггинг и/или бустинг, необходимо рассмотреть минимум 2 алгоритма) для улучшения качества регрессии.
40. Исследование методов генерации синтетических данных для обучения моделей.
41. Исследование методов анализа и визуализации больших объемов данных с использованием графических процессоров (GPU).
42. Оценка и сравнение методов обнаружения дубликатов в больших наборах данных.
43. Исследование методов обработки данных для улучшения устойчивости моделей к шуму и выбросам.
44. Сравнение и анализ различных методов ансамблирования в задачах классификации (необходимо рассмотреть минимум 2 алгоритма).
45. Оценка и сравнение методов стекинга и блендинга моделей в ансамблях. Решаться может задача классификации или регрессии.
46. Исследование и применение методов интеллектуального управления данными.

В контрольной работе представлены следующие разделы:

1. Обзорная часть. Описание метода: (задача, для решения которой разработан метод, основные формулы, алгоритм применения метода, требования к структуре и типу исходных данных, ограничения (условия использования) метода, достоинства, недостатки, …).
2. Основная часть.

- постановка задачи, для решения которой будет использоваться рассматриваемый метод;
- исходные данные (данные должны быть реальными, взятыми из репозиториев данных по машинному обучению, например, ресурс Kaggle, репозиторий UCI (UCI Machine Learning Repository), либо собственные данные).
- описание хода решения задачи в R (привести код на R с комментариями и краткими пояснениями);
- выводы по результатам решения задачи в терминах предметной области (интерпретация результатов).

3. Выводы по работе в произвольной форме (сравнение R с другими средствами анализа данных, оценка рассматриваемого метода анализа данных, анализ результатов и т.п.).

Допускается выбор собственной темы и программного обеспечения анализа данных по согласованию с преподавателем.