

Задание на практические работы

по курсу «Программные системы статистического анализа»

1. Реализация алгоритма на языке R или Python:
 - код с комментариями,
 - если используются пакеты, то подробно прокомментировать используемые входные параметры и выходные значения.
2. Модельный пример.

Для методов машинного обучения описать модель (простой пример: $y = a + bx + \varepsilon$), из каких распределений взяты случайные величины (y , x , ε), истинные значения параметров (a , b).

Для анализа текстов описать, каким образом генерируется случайный текст.

Для анализа графики описать, как генерируются значения для построения графиков.

Ключевая задача – смоделировать **разные** условия эксперимента:

 - разный объем выборки (50, 100, 500, 1000),
 - разное число признаков (единицы, десятки, сотни),
 - разная доля выбросов (5, 10, 20%),
 - разная структура модели (линейная, полином) и т.п.
3. Реальный пример: реальные данные можно взять из курсов Статистический анализ нечисловых данных, Программные системы статистического анализа в диспейс, из репозиторий, для временных рядов можно взять курсы валют и т.п.
4. Исследования на модельных и реальных примерах.

Выходные показатели:

для любых задач: время вычислений,

объем памяти,

для графики: время отображения графика,

размер сохраненного графического файла,

для веб скрапинга и анализа текстов: время загрузки файла,

объем текста,

для задач машинного обучения: время обучения,

время предсказания,

ошибка предсказания,

отклонение оценок от истинных значений параметров.

Все выходные показатели считаются для разных условий эксперименты (см. пункт 2).
5. Выводы, интерпретация: сравнить результаты работы алгоритмов при разных условиях экспериментов, дать заключение, о том какие алгоритмы (пакеты) в каких условиях лучше (хуже) работают.
6. Оформить отчет по ГОСТ.
7. Сравнение с другими языками: провести исследование алгоритма на другом языке (R, Python, C#, Matlab и др.) на **тех же самых** модельных и реальных примерах. Сравнить результаты с выводами из пункта 5.