

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Косынов Сергей Михайлович
Должность: ректор
Дата подписания: 10.06.2024 08:24:34
Уникальный программный ключ:
e3a68f31c62c74154f49980891746bdfcf836

Оценочные материалы для промежуточной аттестации по дисциплине

Основы машинного обучения, 3 семестр

| | |
|--------------------------|---|
| Код, направление | 11.04.02. Инфокоммуникационные технологии и системы связи |
| подготовки | системы связи |
| Направленность (профиль) | Корпоративные инфокоммуникационные системы и сети |
| Форма обучения | Очная |
| Кафедра-разработчик | Радиоэлектроники и электроэнергетики |
| Выпускающая кафедра | Радиоэлектроники и электроэнергетики |

Задание для контрольной работы.

Контрольная работа проводится в формате хакатона по определению злокачественности опухолей на площадке kaggle.

Задание:

1. Зарегистрироваться на kaggle.
2. Цель соревнования: определить, злокачественная опухоль или нет.
3. Целевая переменная – Category, = злокачественная, — незлокачественная. Данные находятся на вкладке Data. Чтобы их скачать, нужно нажать кнопку Download all.
4. На основе данных из train.csv нужно обучить модель машинного обучения и сделать предсказание на основе данных из test.csv.
5. Сформировать файл с решением и отправить его на kaggle. Файл должен выглядеть по структуре так же, как файл submission_sample.csv из той же вкладки Data.

Вопросы к зачету:

1. Смещение (bias) и разброс (variance).
2. Принципиальное отличие байесовского и классического подхода к статистике
3. Методология CRISP-DM
4. Интерпретация моделей машинного обучения
5. Деревья решений и CART алгоритм
6. “Случайный лес” (Random Forest)
7. Градиентный бустинг
8. Гребневая регрессия
9. Логистическая регрессия
10. LASSO
11. Регуляризация
12. Выявление аномалий
13. Алгоритм классификации kNN
14. Метрики качества регрессии MAE RMSE MAPE
15. Метрики качества классификации перекрестная энтропия, энтропия, ассигасу, точность, полнота, F-мера.
16. Алгоритмы снижения размерности
17. Алгоритмы кластеризации
18. Ошибки первого и второго рода, уровень значимости и мощность
19. p-value - что означает и как интерпретировать
20. Процедура кросс-валидации

21. Кросс-валидация для временных рядов
22. Решение проблемы дисбаланса классов
23. Решение проблемы пропущенных значений
24. Кривые валидации и обучения.
25. Стекинг
26. Кодирование категориальных признаков
27. Биннинг признаков
28. Отбор признаков
29. Извлечение признаков из текста
30. Стандартизация признаков
31. Проблемы, вызванные скоррелированными признаками