**Название Курсовой: Решение задачи … с применением методов Data Mining**

1. Выбираете задачу, которую хотите решить и класс задач МО (классификация, регрессия, предиктивная аналитика). Сетевые атаки, потребление электроэнергии, распознавание образов, текста, аудио, кластеризация данных, etc. Придумайте задачу, которую вам будет интересно решить.
2. Выбираете датасет. Не стесняйтесь ходить за вдохновением на Kaggle.com
3. Изучаете датасет. Какие колонки, где есть пропуски, какое распределение, построение гистограммы. Визуализация классов.
4. Ищите выбросы
5. Предобработка датасета (Изучение количества пропусков, принятие решение о стратегии их обработки), изменение типов переменных в удобоваримые для МО
6. Построение матрицы корреляции. Сделать вывод о степени зависимости переменных. Если есть сильная зависимость – удалить столбец
7. Выбор 2+ алгоритмов МО для построения модели. Пишите теоретическую выкладку по каждому: принцип работы, какие параметры, какие задачи решает (лучше с формулами)
8. Разбиение датасета на обучающую и тестовую выборку
9. Выбор параметров для обучения (через Gini или другим способом)
10. Обучение алгоритмов МО (с stratified k-fold, предварительно описать, что это такое)
11. Сделать вывод о скорости обучения
12. Визуализация результатов. ROC-AUC, Accuracy, F1-score и другие метрики качества, которые изучали.
13. Вывод об эффективности моделей

25+страниц

Порядок пунктов можно немного изменить

Алгоритмы, которые можете применять: NB,LR,SVM,KNN,DT,RF,GB(любая из реализаций). Можно в качестве второго алгоритма использовать нейронки.