

Промежуточная аттестация

МОДУЛЬ 2 «Нейронные сети и алгоритмы глубокого обучения»

Формулировка задания

Улучшить результат выбранной модели с помощью feature engineering и оценки важности признаков из алгоритма случайного леса.

Планируемый результат

Улучшен результат обучения выбранной модели с помощью внедрения новых признаков.

Описания плана работы

- 1) Выбрать модель из ранее обученных на курсе. Обратить внимание, что модель должна содержать, в основном, числовые столбцы (данные с погодой, данные для алгоритмов knn, svm, бустинга).
- 2) Использовать блокнот с выбранной обученной моделью.
- 3) Сгенерировать новые признаки с помощью библиотек tsfresh, featuretools или feature-engine. Обратить внимание, что все столбцы с максимальным числом значений 0 или NaN необходимо исключить из новых признаков.
- 5) Выбрать новые признаки с наибольшей важностью для дополнения исходных данных. Использовать оценку важности

признаков из алгоритма случайного леса (RandomForestRegressor или RandomForestClassifier).

- 6) Дополнить исходные данные по результатам п.5. Обратить внимание, что дополнить данные нужно небольшим количеством наиболее важных признаков.
- 7) Обучить ещё раз модель
- 8) Представить график кривой обучения
- 9) Сравнить результаты обучения модели с предыдущими показателями. Улучшилось ли качество результата?
- 10) Провести эксперимент с добавлением разного количества новых столбцов. В каком случае получен максимальный результат по метрикам качества для текущего алгоритма?