# Результаты, сгруппированные по методам машинного обучения

Для исследования было выбрано большое количество методов машинного обучения, которые были сгруппированы в следующие категории:

* Boosting
* Clustering
* Regression
* Decision trees
* Nearest neighbours
* Support vector machine
* Hyper pipes (heuristic)
* Voting Feature Intervals (heuristic)
* Meta classifiers (weak classifiers which used by another classifiers)

# Результаты, сгруппированные с учетом дисперсии

Все результаты отображены в виде графиков на одной диаграмме:

Диаграмма 1

Как можно заметить, графики сильно накладываются друг на друга, ввиду чего далее представлены диаграммы, на которых изображены вместе отдельные графики из первой диаграммы, сгруппированные по определённым отличительным признакам.

Диаграмма 2

Диаграмма 2 хорошо отражает схожесть результатов обучения алгоритмов, основанных на деревьях решений, регрессии и VFI – несмотря на то, что верхняя граница дисперсии этих методов расположена очень высоко, эти графики имеют очень большую дисперсию, а также большую площадь перекрытия.

Диаграмма 3

Диаграмма 3 наглядно иллюстрирует сходство самых малодисперсных методов – мета-классификаторов и HyperPipes. Несмотря на их малую дисперсию, по сравнению с другими группами методов, эта группа методов имеет меньшее значение NCG, следствием чего являются более низко расположенные графики.

Диаграмма 4

На диаграмме 4 отображена группа методов, основанная на кластеризации. Этот метод проявляет большую дисперсию и имеет достаточно низкое, по сравнению с другими группами, положение «средней» линии. Особый интерес вызывает то, что для небольшого количества постов (до 90) эта группа методов имеет крайне малую дисперсию.

Диаграмма 5

Как видно из 5 диаграммы, группа SVM методов выдает отличные результаты на небольшом количестве постов (до 50), после чего результативность несколько снижается, однако несмотря на это, имеет малую дисперсию. Ключевым отличием от далее рассматриваемой группы является характерная немонотонность графика NCG.

Диаграмма 6

На последней диаграмме отражены группы методов, которые выдали наилучшие результаты по сравнению с остальными. Группы методов, основанных на бустинге и методов ближайших соседей, имеют монотонный график NCG, высокое положение линии «среднего» и небольшую дисперсию.

# Результаты, сгруппированные только с учетом линии среднего

Все результаты отображены в виде графиков на одной диаграмме:

Диаграмма 1

Далее представлены диаграммы, на которых изображены вместе отдельные графики из первой диаграммы, сгруппированные по определённым отличительным признакам.

Диаграмма 2

Диаграмма 2 наглядно иллюстрирует сходство методов мета-классификаторов и HyperPipes. Эти методы почти повторяют свою линию «среднего».

Диаграмма 3

На диаграмме 3 отображена группа методов, основанная на кластеризации. Этот метод имеет достаточно низкое, по сравнению с другими группами, положение «средней» линии.

Диаграмма 4

Как видно из 4 диаграммы, группа SVM методов выдает отличные результаты на небольшом количестве постов (до 50), после чего результативность несколько снижается. Ключевым отличием от далее рассматриваемой группы является характерная немонотонность графика NCG.

Диаграмма 5

На 5 диаграмме отражены группы методов, которые выдали наилучшие результаты по сравнению с остальными. Эти графики имеют очень похожие линии «среднего». В отличие от группировки с учетом дисперсии, в эту группу регрессия и бустинг входят одновременно.

Диаграмма 6

На последней диаграмме расположена группа методов, основанная на VFI. При учете дисперсии этот метод входил в группу методов с большой дисперсией. При рассмотрении линии «среднего» в отдельности хорошо видно, что она сильно отличается от всех других методов.