*Краткий отчет по статье: “Ensembled Boosted Trees with Synthetic Features Generation in Application to Bankruptcy prediction”*

**1. Проблема и цель статьи.**

Прогнозирование банкротства предприятия с целью уменьшения социального и экономического риска.

Цель: разработать модель, которая включает в себя различные эконометрические измерения (**синтетические показатели**) и позволяет прогнозировать финансовое состояние фирмы.

**2. Проблемы при составлении модели, подбора данных**

Исследователи из данной статьи приводят 2 проблемы связанные с прогнозированием банкротства:

**Проблема 1**: Экономические индикаторы, описывающие состояние фирмы, устанавливаются экспертами в области экономики. (Ни сразу может быть понятно, как интегрировать показатели в 1 модели)

**Проблема 2**: Данные для прогнозирование обычно несбалансированные. (Компаний-банкротов намного меньше, чем предприятий, которые успешно работают)

Обе эти проблемы влияют на конечный прогноз.

**3. Методология**

В основе статьи лежат такие понятия: алгоритм extreme gradient bosting, генерация синтетических, искусственных показателей, ensemble of boosted trees.

**Синтетические показатели –** этоискусственно созданные индикаторы, которые сгенерированы с помощью 2 случайных показателей из основного набора данных и 4 случайных арифметических операций (+,-,/,\*).

**4. Данные**

В каждом наборе данных содержатся 64 финансовых показателя + класс прогнозирования (0 – не банкрот, 1 – банкрот)

1st year – данные содержат финансовые показатели с 1-го года прогнозируемого периода и метку соответствующего класса, которая указывает на статус банкротства через 5 лет.

2nd year – данные содержат финансовые показатели за 2-й год прогнозируемого периода и метку соответствующего класса, которая указывает на состояние банкротства через 4 года.

3rd year – данные содержат финансовые показатели за 3-ий год прогнозируемого периода и метку соответствующего класса, которая указывает на состояние банкротства через 3 года.

4th year – данные содержат финансовые показатели за 4-й год прогнозируемого периода и метку соответствующего класса, которая указывает на состояние банкротства через 2 года.

5th year – данные содержат финансовые показатели за 5-й год прогнозируемого периода и метку соответствующего класса, которая указывает на состояние банкротства через 1 год.

**5. Эксперимент**

Для тренировки и выбора оптимальных гиперпараметров модели была выполнены 10-ти фолдовая кросс валидация.

В связи с несбалансированными данными, авторы статьи использовали AUC кривую для оценки качества моделей.

Для ensemble boosted trees авторы получили наилучшие результаты для числа базовых учащихся (base learners), равного 10, и количества синтетических показателей, сгенерированных на каждой итерации, равных 60. Искусственный финансовый показатель принимается для следующей итерации, если он наблюдался как минимум в 5% деревьев в лесу.

**Результаты**

Результаты эксперимента представлены в таблице ниже (рис. 1). Для каждого из рассмотренных периодов и исследованных моделей авторы представили среднее (MN) и стандартное отклонение (STD) для измерения AUC, которое было рассчитано на основе 10ти фолдовой кросс валидации.

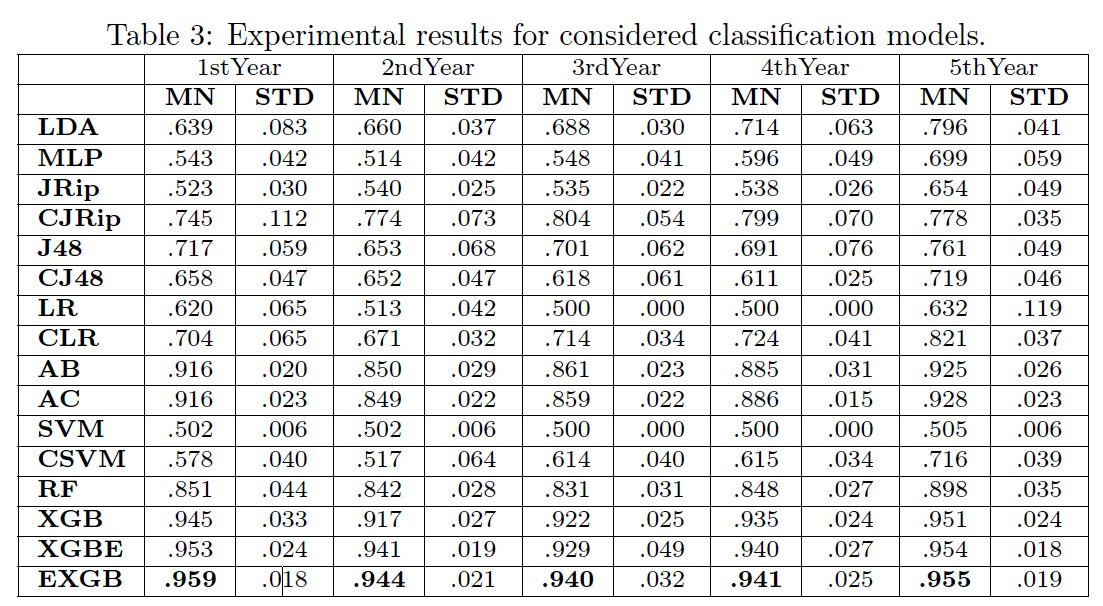


Рисунок 1 – результаты эксперимента для рассматриваемых классификационных моделей

**Анализ важности показателей**

Авторы статьи оценили важность признаков, рассчитав их общее количество, наблюдаемое в узлах структуры леса, по общему количеству узлов в деревьях, составляющих лес. Другими словами, они приняли во внимание категориальное распределение.

Проанализировав результат, исследователи выделили следующие важные финансовые показатели:

X25 – скорректированная доля собственного капитала в финансовых активах;

X40 – Коэффициент текущей ликвидности;

X52 – Коэффициент оборачиваемсти кредиторской задолженности.

Также выделяют следующие индикаторы, которые тоже могут быть полезными в прогнозировании.

X13, X22, X31, X42 – коэффициенты прибыли;

X15 – коэффициенты кредитного плеча;

X9, X36, X48, X52 – показатели операционной эффективности;

X5, X27, X58 – иные показатели.