

## Введение

Одним из этапов исследования данных на сегодняшний день является визуальный анализ. Он проводится с помощью различного рода диаграмм и графиков. Такой анализ позволяет выявлять закономерности, находить аномалии, и многое другое. Также после введения модели в эксплуатацию одним из важнейших для поддержания ее работы процессов является мониторинг показателей.

В качестве инструмента для визуального анализа и мониторинга показателей часто используется язык программирования с соответствующими библиотеками, например, язык Python с пакетом `seaborn`. Очевидным преимуществом такого подхода является высокая гибкость, но расплачиваться за это приходится достаточно большим количеством времени, необходимого для создания информативного графика и внесения даже минорных изменений.

Помимо описанного выше подхода, можно использовать специализированное ПО. Оно позволяет снизить порог вхождения благодаря отсутствию необходимости в знании программного API, упростить создание и изменение диаграмм с помощью графического интерфейса, а также решает большое количество неявных проблем. Но в все эти преимущества идут рука об руку с весомыми ограничениями. Например, поддержка только определенного формата, всего несколько возможных типов графиков, дополнительные ограничения, вызванные интерфейсом.

В моей работе предлагается совместить оба подхода для создания программного решения, обладающего изложенными выше плюсами и лишенного если не всех, то хотя бы нескольких недостатков, с целью создать интерактивную панель с диаграммами, позволяющую обеспечить простой и удобный мониторинг определенных показателей кредитного портфеля.

## 1 Постановка задачи

Целью работы является с помощью разработанного программного решения сконструировать интерактивный дашборд – панель с диаграммами. Ниже представлены требования как для ПО, так и для дашборда.

В качестве источника данных изначально должен поддерживаться формат csv как наиболее распространенный, но дальнейшее добавление новых источников должно быть простым и понятным. Должна существовать возможность разместить панель диаграмм на сервере, к которому можно подключаться с удаленного клиента.

Все диаграммы дашборда показателей должны поддерживать фильтрацию по дате, типу кредита, продукту, флагу дефолта и причине дефолта. Диаграммы должны отображать следующие показатели:

- Численность договоров – количество договоров на дату с учётом всех фильтров в виде линейного графика
- Численность клиентов – количество клиентов на дату с учётом всех фильтров в виде линейного графика
- Default rate – доля тех, кто выйдет в дефолт в течение года, от тех, кто не в дефолте на дату оценки в виде линейного графика
- Уровень модельных оценок PD – среднее значение по полю вероятности дефолта с учётом всех фильтров. Исторический период – в виде линейного графика, а также **барометр** для самой последней даты.
- Распределение по рейтингам - столбчатая диаграмма
- Качество модели - показатель Gini.

## 2 Разработка платформы конструирования дашборда

### 2.1 Поиск и сравнение существующих решений

Одним из ключевых атрибутов системы должна быть открытость, т.е. это должно быть open source решение. Это позволит добиться той же гибкости, что дает использование только языка программирования. Из-за этого сразу отпадают лучшие приложения для исследования и визуализации данных: Qlik и Tableau.

Среди open source платформ наиболее популярная и удобная это Grafana. Она позволяет создавать интерактивные дашборды с большим набором возможных типов диаграмм и источников данных, а плагины сообщества дополнительно расширяют функционал. Главным недостатком с точки зрения данной работы является отсутствие поддержки в качестве источника данных внешнего файл – все предлагаемые источники — это различные базы данных. Но Grafana обладает достаточно понятным API, что позволяет самому создать плагин. Именно это стало основополагающим фактором в выборе платформы конструирования дашборда.