



V ВСЕРОССИЙСКАЯ НАУЧНО-ПРАКТИЧЕСКАЯ КОНФЕРЕНЦИЯ
Анализ и прогнозирование развития экономики России
Сергиев Посад, 23-24 марта 2023

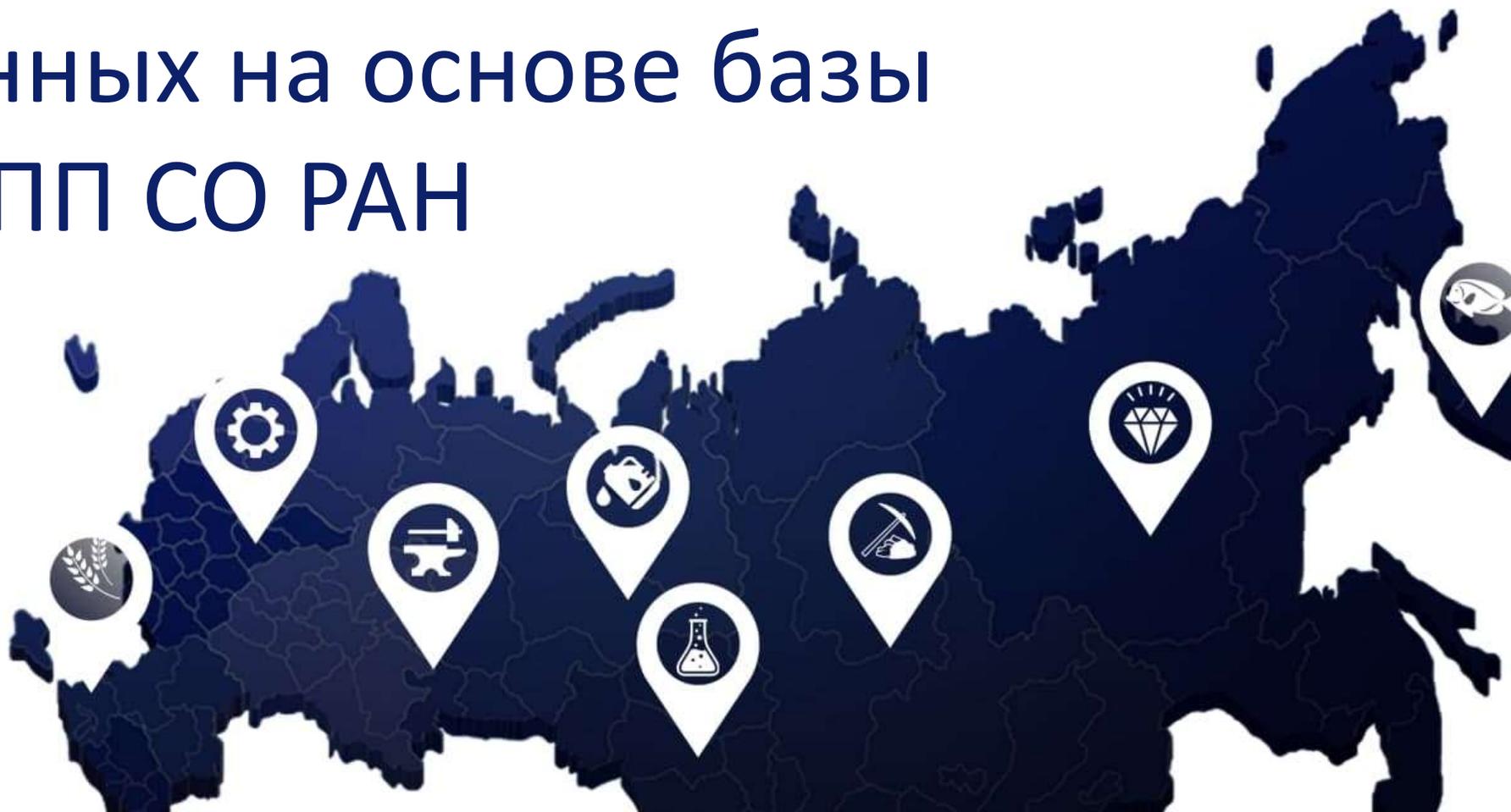
Формирование системы анализа больших данных на основе базы знаний ИЭОПП СО РАН

Костин Андрей

с.н.с, к.э.н.

Родионова Дарья

Инженер



Система поиска и коррекции данных с низким уровнем достоверности





Несоответствие показателей
отчетности



Всплески в показателях
(напр, в показателе выручки)

Модуль 1. Всплески

Оценка темпов прироста выручки
компании

Модуль 2. Финансовые показатели

Проверка всех связанных показателей
финансовой отчетности компании

Ошибки	Доля в %
Без ошибок	94,96%
Ошибки баланса	4.46%
Всплески	0.52%
Ошибки баланса + Всплески	0.06%

Возникает задача кластеризации



Общее описание показателей

199 МБ

- более 4 тыс. показателей

- 96 регионов

Список подпараметров

4,2 МБ

Данные: панельные данные

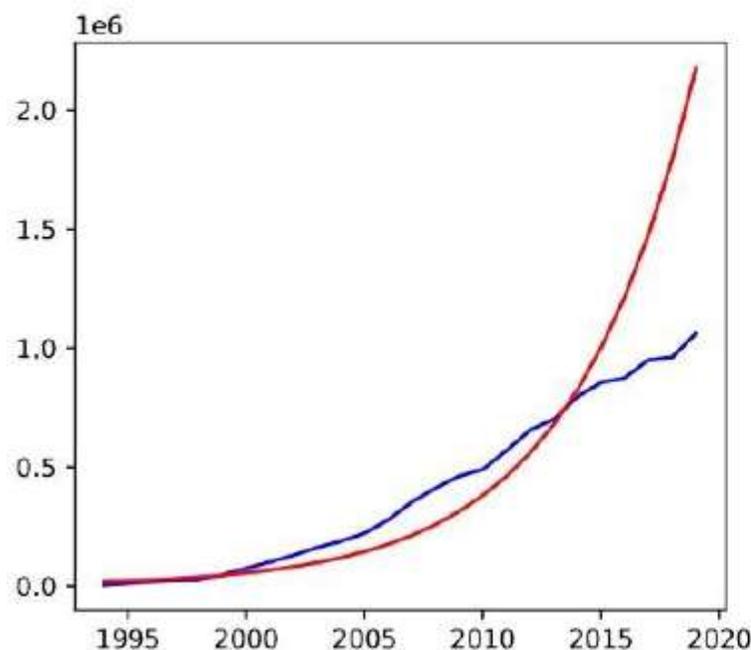


Анализ гетероскедастичности

Аналитика для каждого из показателей

Тесты

Гольфельда-Квандта и Брейша-Пагана



Номер параметра: 637

Пропуски в данных: не обнаружены

Количество наблюдений: 26

Единицы измерения: млн. рублей

R2: 0.901

Описание: валовая добавленная стоимость по видам экономической деятельности и секторам (в текущих ценах; миллионов рублей) в соответствии с общероссийским классификатором видов экономической деятельности ок 029-2014 оквэд2, сектор "некоммерческие организации, обслуживающие домашние хозяйства": добыча полезных ископаемых

Результаты теста на гетероскедастичность:

Тест	95 %	90 %
Гольфельд-Квандт	1	1
Брейш-Паган	1	1

Данные: панельные данные



Анализ стационарности

Тест Дики-Фуллера

Анализ коинтеграции

Тест Педрони

Процент стационарных данных по группам параметров



Прогнозирование:

- ARIMA, ADL модели,
- деревья решений,
- причинность по Грейнджеру
- LSTM модели

Таможенная статистика

Выберите Годы:



Выберите регионы:

РФ

Выберете параметр:

- Импорт
- Экспорт
- Экспорт и Импорт

Выберете параметр ТН ВЭД:

- Все
- Выбрать по Кодам

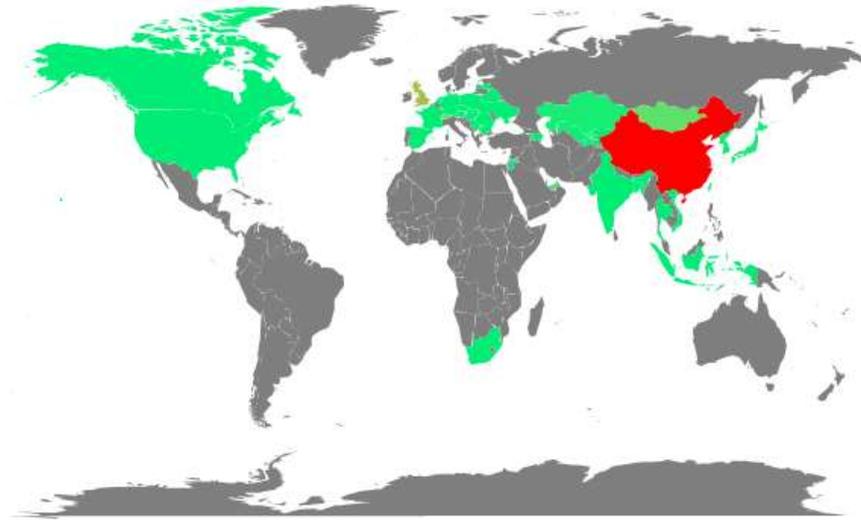
Загрузить xlsx

Скачивание может занять достаточно долгое время. Если файл будет большим, то нам придется разбить его на несколько частей. Вам придется нажать кнопку Загрузить xlsx еще раз после скачивания файла

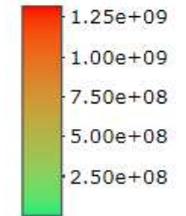
Скачать каталог кодов ТН ВЭД

Построить график суммарных потоков:

Суммарные объемы



value

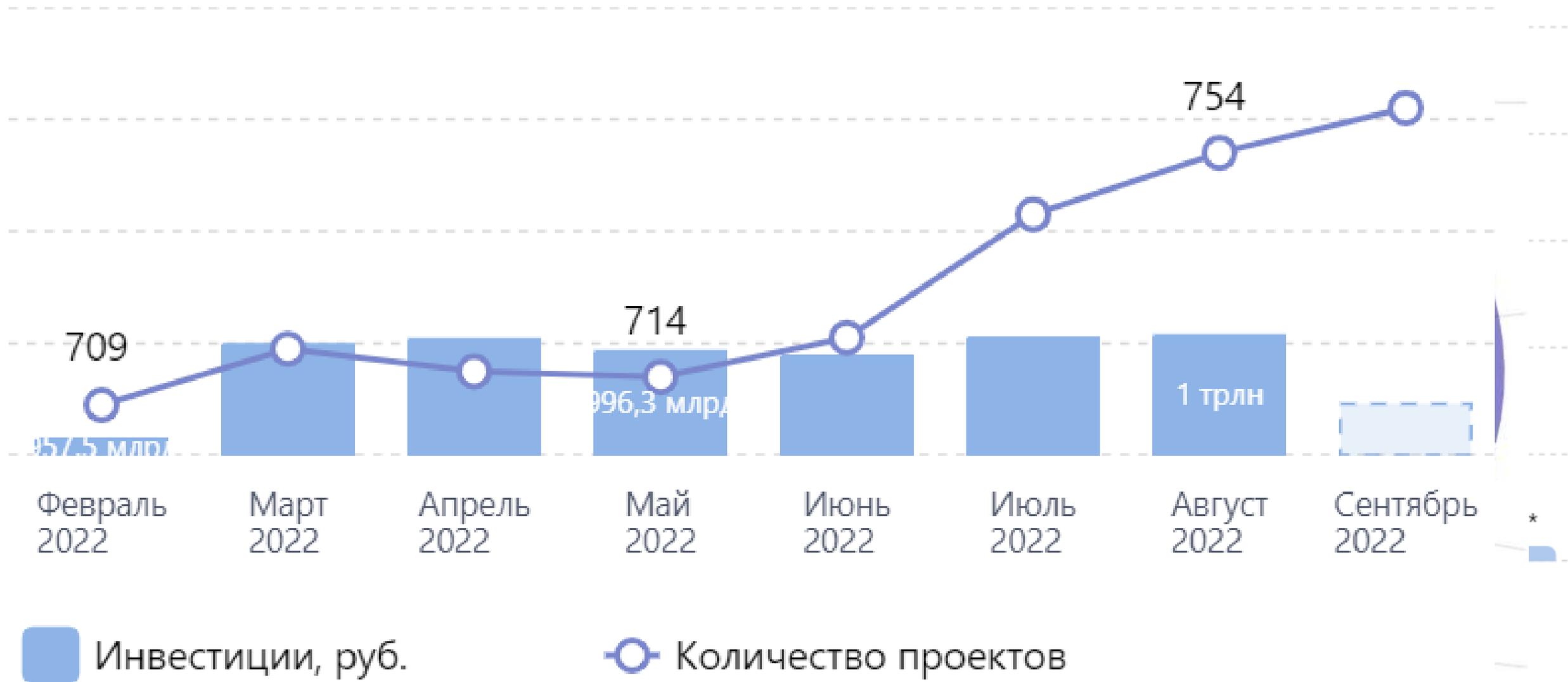


355

741

535

318





Спасибо!

a.kostin@nsu.ru