

Ансамблевые алгоритмы для оценки основных макроэкономических показателей

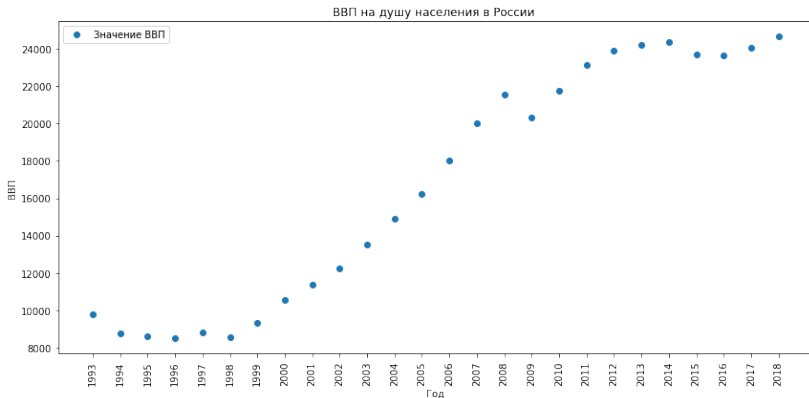
Рогоза Ярослав

Московский Государственный Университет имени М. В. Ломоносова
Экономический факультет

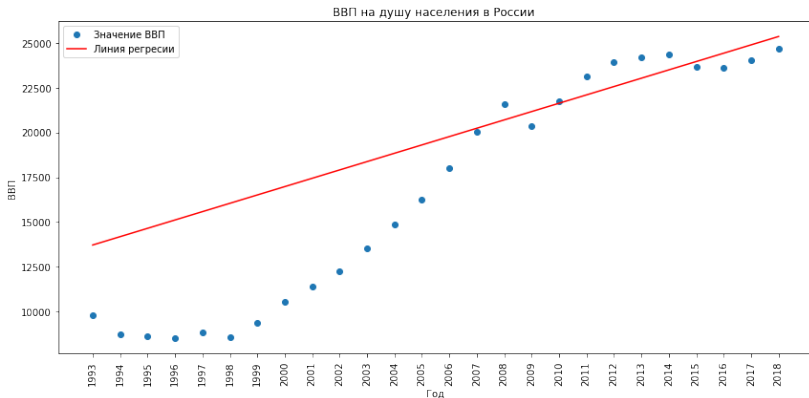
Предмет: Макроэкономика-2
Группа 208

2023

С помощью чего, по Вашему мнению, можно предсказать значение ВВП в России в 2019 году?



Самый простой способ составить уравнение линейной регрессии. Но почему этот результат будет плохим?



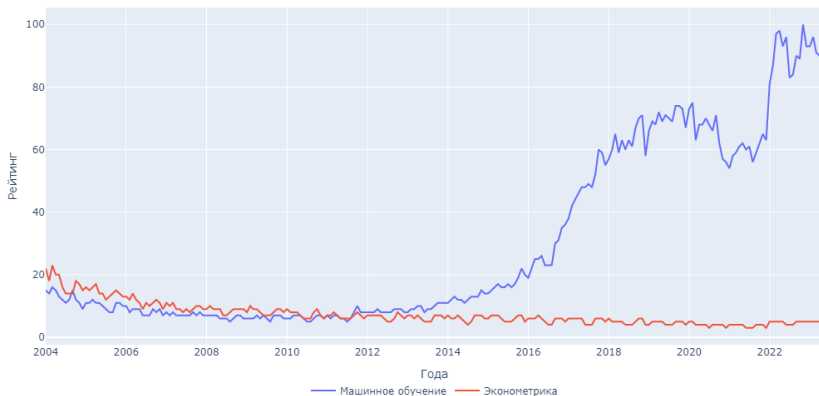
Другие способы оценки:

- ▶ LASSO, RIDGE - регрессии
- ▶ VAR, ARIMA и их комбинации с другими методами
- ▶ Нейросети
- ▶ Машинное обучение

Актуальность темы

В 2011 году по рейтингу гугла по количеству ежедневных запросов машинное обучение и эконометрика сравнялись, после чего начался резкий рост машинного обучения

Рейтинг гугла по количеству ежедневных запросов



Публикации в журнале Банка России

- ▶ Шуляк Е. Макроэкономическое прогнозирование с использованием данных социальных сетей - 2022 //
 - ▶ Джункеев У. Прогнозирование безработицы в России с помощью методов машинного обучения – 2022 //
 - ▶ Семитуркин О. Шевелев А. Корректное сравнение предиктивных свойств моделей машинного обучения на примере прогнозирования инфляции в Сибири - 2023 //
- Вывод: "Прогнозирование методом комбинирования моделей машинного обучения в большинстве случаев предпочтительнее, чем с использованием какой-то одной модели."

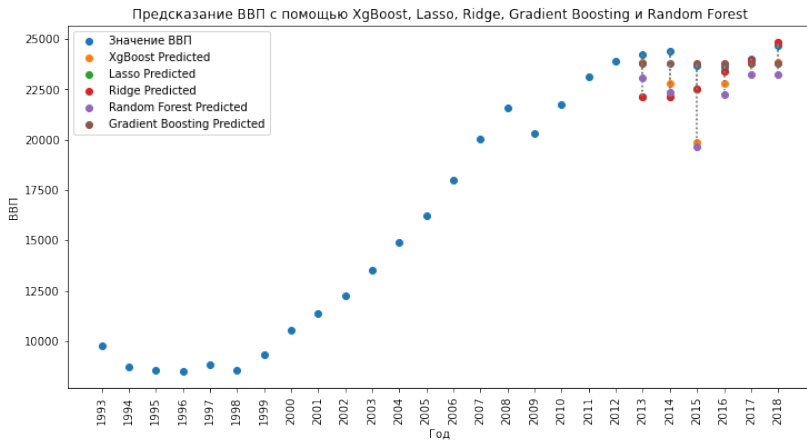
Постановка проблемы

Действительно ли методы машинного обучения, а в частности ансамблевые алгоритмы справляются с задачей оценки макроэкономических показателей лучше, чем классические эконометрические модели?

Условия эксперимента:

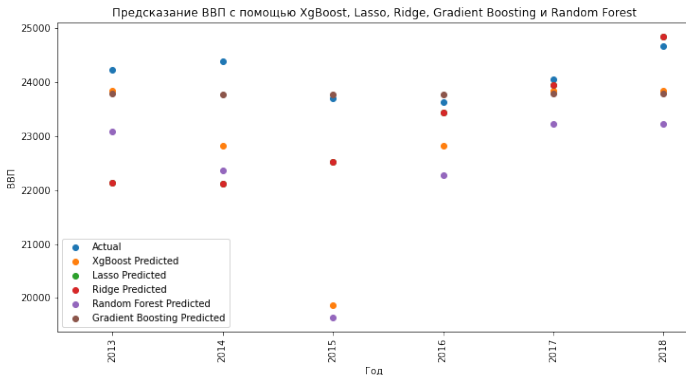
- ▶ Были использованы следующие модели: XgBoost, Lasso, Ridge, Gradient Boosting и Random Forest
- ▶ Обучение моделей происходило на данных ВВП на душу населения в России с 1993 по 2012 года или же 80% от всей выборки
- ▶ В качестве дополнительных факторов, влияющих на ВВП были выбраны объем производство газа, нефти, а также показатели инфляции и безработицы
- ▶ В качестве метрики для сравнения было использовано среднеквадратическое отклонение RMSE

Эксперимент (80% обучающая выборка)



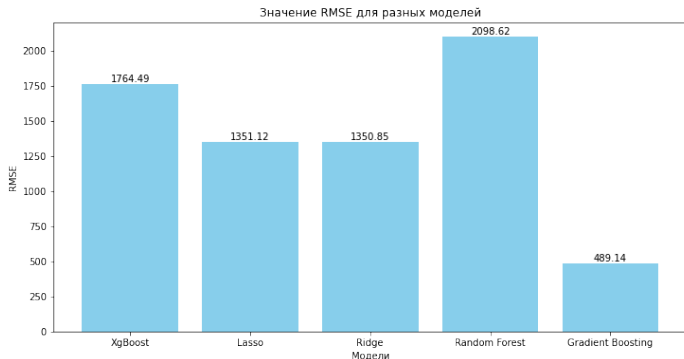
Эксперимент (80% обучающая выборка)

Увеличенный предыдущий график:



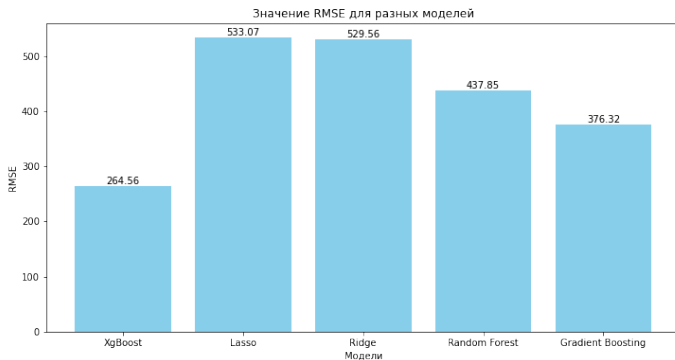
Эксперимент (80% обучающая выборка)

Среднеквадратическое отклонение прогноза каждой модели:



Эксперимент (90% обучающая выборка)

Если мы возьмем 90% данных в качестве обучающей выборки и будем прогнозировать ВВП на 3 года вперед, то получим следующий результат:



Сравнение авторегрессий с ансамблевыми алгоритмами

В данной таблице показаны данные полученные на основе прогнозных значений ИПЦ в Сибири. Среднеквадратическое отклонение ARIMA является базовым показателем, на который были поделены RMSE остальных моделей (чем меньше значение, тем лучше модель)

Метод	Горизонт прогнозирования, месяцы				
	3	6	9	12	24
Градиентный бустинг	0,87	0,76	0,98	0,70	0,76
Комбинированный прогноз	0,90	0,85	0,91	0,57	0,79
Случайный лес	0,78	0,78	0,93	0,57	0,86
Гребневая регрессия	1,12	1,03	1,06	0,66	0,88
Эластичная сеть	1,10	1,09	0,99	0,65	0,90
LASSO	0,84	0,83	0,89	0,56	0,91
ARIMAX	0,69	0,61	0,69	0,58	0,95
ARIMA	1,00	1,00	1,00	1,00	1,00

Минусы использования ансамблевых алгоритмов в оценке макроэкономических показателей

- ▶ Высокая вычислительная сложность
- ▶ Ансамблевые модели обычно более сложны в интерпретации по сравнению с простыми линейными моделями
- ▶ Возможно переобучение

Плюсы использования ансамблевых алгоритмов в оценке макроэкономических показателей

- ▶ Более точные прогнозы по сравнению с классическими эконометрическими моделями
- ▶ Устойчивость к выбросам, шуму, мультиколлинеарности
- ▶ Гибкость и разнообразие моделей

Спасибо за внимание!