

Использование метода экстремального градиентного бустинга в прогнозировании консолидированного бюджета региона

Using the Method of Extreme Gradient Boosting in Forecasting the Consolidated Budget of the Region (DOI: 10.34773/EU.2021.5.23)

Б. СУЛТАНОВ, Д. СУЛТАНГАРЕЕВ

Султанов Булат Рамдисович, аспирант кафедры цифровой экономики и коммуникации в экономике Института экономики, финансов и бизнеса Башкирского государственного университета (ИНЭФБ БашГУ). E-mail: sultanoff.bulat.3.4@yandex.ru

Султангареев Даниил Рамильевич, аспирант кафедры социологии труда и экономики предпринимательства ИНЭФБ БашГУ. E-mail: spacesuminchd@gmail.com

В статье рассматривается использование метода экстремального градиентного бустинга, позволяющего прогнозировать объемы основных доходных источников бюджета на среднесрочный и долгосрочный периоды. Проанализированы данные прогноза социально-экономического развития и сформировавшихся тенденций социально-экономического развития за ряд предыдущих лет. Доказано, что посредством включения в систему дополнительных ограничений по объему государственного долга и дефицита бюджета, установленных бюджетным законодательством, аналитическая система позволяет прогнозировать объем расходов и, соответственно, дефицита бюджета.

Ключевые слова: прогнозирование, ансамбль регрессионных моделей, показатели социально-экономического развития, алгоритмы машинного обучения, экстремальный градиентный бустинг.

The article discusses the use of the method of extreme gradient boosting, which is able to predict the volume of the main revenue sources of the budget for the medium and long term. Analyzed the data of the forecast of socio-economic development and formed trends in socio-economic development for a number of previous years. It is proved that by including in the system additional restrictions on the volume of public debt and budget deficit established by the budget legislation, the analytical system makes it possible to predict the volume of expenditures and, accordingly, the budget deficit.

Key words: forecasting, ensemble of regression models, indicators of socio-economic development, machine learning algorithms, extreme gradient boosting.

Показатели социально-экономического развития ежегодно не достигаются или перевыполняются, затрудняя адекватное планирование и исполнение бюджета, что является основной проблемой при разработке его проекта для соответствующей территории (далее – региона, регион). Второстепенной проблемой является ежегодное неоднократное (более 7 раз в год) внесение изменений в закон о бюджете, при этом отдельные изменения обусловлены изменениями прогноза социально-экономического развития.

Данные проблемы влекут за собой невозможность достижения установленных бюджетом показателей, так как планирование бюджета основывается на прогнозе социально-экономического развития региона. При этом существующие способы и методы моделирования показателей регионального бюджета не позволяют в полной мере достоверно определить прогнозные значения в зависимости от изменений показателей прогноза социально-экономического развития, в том числе в связи с недостижением (перевыполнением) отдельных параметров прогноза.

Методика решения данной задачи должна включать в себя следующие входные данные:

1. Прогнозы социально-экономического развития региона на среднесрочный период и отчеты об исполнении основных параметров прогноза социально-экономического развития региона на среднесрочный период;
2. Месячные отчеты об исполнении бюджета региона за январь;
3. Годовые отчеты об исполнении бюджета региона.

С учетом вышеописанного перечня данных, нами предлагается разработать аналитическую систему для прогнозирования основных параметров регионального бюджета, исходя из сложившихся трендов на основе ретроспективного анализа прогнозов социально-экономического развития и их исполнения во взаимосвязи с фактическим исполнением бюджета на примере субъекта Российской Федерации – Республики Карелия.

Чтобы реализовать данный метод, необходимо выделить компоненты финансов региона. Так, по мнению Р.А. Попова (с которым мы согласны): «Финансы современного российского региона представляют собой систему отношений по поводу формирования денежных средств территории и их использования» [1]. По его же мнению, финансы региона состоят из [1]:

- регионального бюджета;
- региональных внебюджетных фондов.

Они формируются (пополняются) за счет нормированной части федеральных налогов и сборов, а также исключительно региональных налогов и сборов, штрафов, займов, лотерей и т.п. [1], как продемонстрировано на рис. 1.

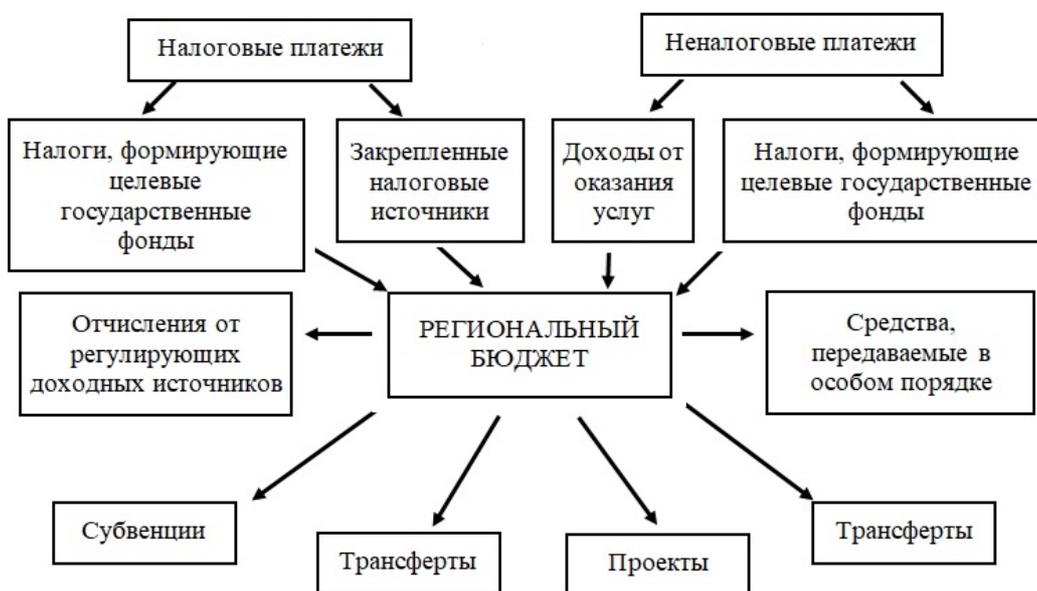


Рис. 1. Модель формирования регионального бюджета

При использовании метода прогнозирования для консолидированного бюджета региона необходимо отметить, что консолидированный бюджет региона – это совокупность регионального и местного бюджетов, включающих в себя местный бюджет и местные внебюджетные фонды – в данном случае, Республики Карелия.

По Р.А. Попову, в планировании регионального бюджета должны использоваться следующие методы [1]:

- 1) методы экспертных оценок;
- 2) методы обработки пространственных, временных и пространственно-временных совокупностей;
- 3) методы ситуационного анализа и прогнозирования.

Для планирования консолидированного бюджета Республики Карелия повышенный интерес представляет метод ситуационного анализа и прогнозирования. В основе данного метода находятся модели, предназначенные для анализа функциональных (т.е. жестко детерминированных) зависимостей, в которых каждому значению фактурного признака соответствует значение результативного признака, формой которого является региональный бюджет на год, где результативный признак «расходы» имеет сильную корреляцию с фактором «доходы» [1].

Методы

Прогнозирование общих доходов Республики Карелия выполнено в информационной программно-аналитической среде RStudio, представляющей эффективные алгоритмы анализа временных рядов при помощи различных подходов машинного обучения.

Исследование исходного социально-экономического показателя «Доходы бюджета – всего» проведено в пять этапов.

На первом этапе осуществлена предварительная обработка базы данных, структурирование, нормирование и центрирование переменных.

На втором этапе из многочисленных методов оценки (авторегрессионная модель проинтегрированного скользящего среднего ARIMA, SARIMA, ARIMAX, авторегрессионная модель с распределенными лагами ADL и т.д.) было решено применить метод экстремального градиентного бустинга с отложенным воздействием результирующей переменной (далее – XGBoost).

XGBoost – это метод машинного обучения, позволяющий для увеличения точности предсказания строить модель классификации в виде композиции базовых классификационных моделей, а именно, деревьев решений [2].

XGBoost основан на последовательном построении ансамбля регрессионных моделей, которые с каждой итерацией минимизируют среднеквадратичную ошибку. Выбор в пользу данного подхода обусловлен тем, что этот алгоритм обладает высокой предсказательной способностью, а также включает в себя различные способы регуляризации, что уменьшает переобучение и улучшает общую производительность [3].

На третьем этапе методом цепных уравнений Маркова восстановлены пропущенные значения. Данная необходимость вызвана применением лагированной переменной доходов бюджета. Технология использует комплексные функции для работы с пустыми значениями (рис. 2).

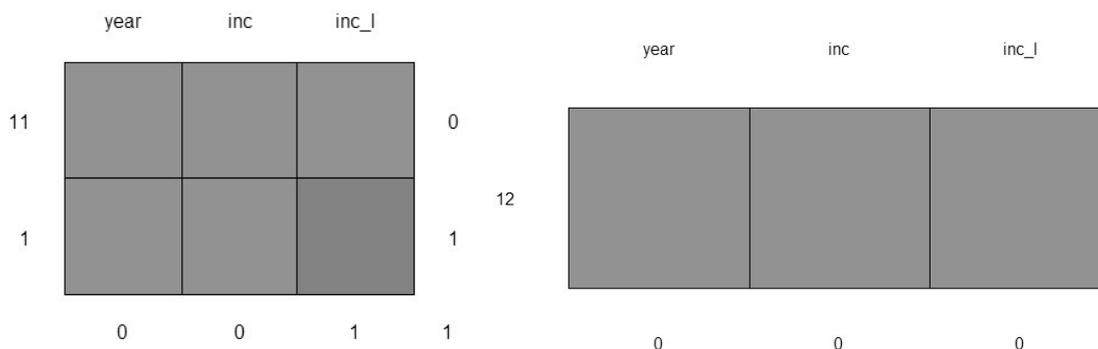


Рис. 2. Структура исходных данных до и после восстановления пропущенных значений

На четвертом этапе проводилось моделирование. Сначала исходная выборка была разделена на обучающую и тестовую в пропорции 4:1. Затем были построены разреженные матрицы выборок. По построенной модели экстремального градиентного бустинга с лаговыми значениями эндогенной переменной проведено прогнозирование.

Результаты

Пятый этап был посвящен проверке прогностической способности модели. Средняя абсолютная процентная ошибка составила 21 %. Соответственно, точность модели – 79 % (рис. 3).

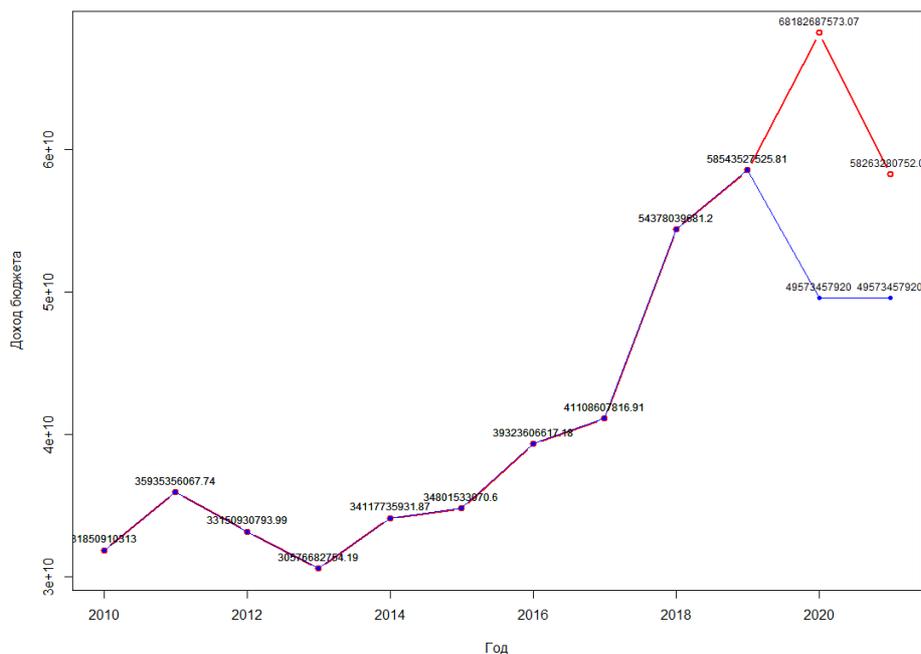


Рис. 3. График фактических и прогнозных значений результирующей переменной

Обсуждение

Результаты работы соответствуют цели и гипотезам исследования. Итоги моделирования позволяют использовать полученную модель в практических и прикладных целях. В дальнейшем планируется провести исследования смежных социально-экономических показателей на расширенном временном промежутке.

Заключение

Таким образом, обобщая вышеизложенное, можно сделать вывод, что применение метода экстремального градиентного бустинга показывает довольно высокую эффективность и прогнозную способность. Благодаря использованию методов машинного обучения появляется возможность учёта множества параметров, влияющих на изменение расходов и доходов бюджета. Применён метод стэкинга – ансамбля моделей для использования преимуществ разных регрессионных моделей.

Литература

1. Региональное управление и территориальное планирование: Учебник / Р.А. Попов. М.: НИЦ ИНФРА-М, 2013. 288 с.
2. Natekin A. and Knoll A. Gradient boosting machines a tutorial // Front. Neurorobot. 2013. Vol. 7. P. 21.
3. Бустинг с помощью AdaBoost и GradientBoosting [Электронный ресурс]. URL: <https://lambda-it.ru/post/busting-s-pomoshchiu-adaboost-i-gradient-boosting>