**Наукастинг элементов использования ВВП России с учётом анализа новостного фона**

Большое число экономических агентов: как органы государственной власти, так и представители частного сектора нуждаются в получении максимально оперативной оценки текущего состояния экономики, что становится особенно актуально в условиях неопределенности. В работе рассматривается вопрос оперативной оценки (наукастинга) текущих темпов роста ВВП России и его компонентов по использованию на квартальных данных. Официальная статистика, например, по ВВП и его компонентам, выходит со значительной задержкой, а далее еще и корректируется, что повышает важность проблемы оперативной оценки макроэкономических показателей по уже имеющимся данным. В современной исследовательской практике для решения данной проблемы используется наукастинг – методика оценки текущих уровней показателей на основе более оперативных данных. Данная работа дополняет библиотеку отечественных работ, посвященных изучению наукастинга на российских данных, являясь при этом одним из первых исследований по построению моделей для наукастинга элементов использования ВВП России, и первым, использующим для этих целей анализ новостного фона. Новостной фон – показатель, доступный исследователю практически в режиме реального времени, а его включение в качестве одного из регрессоров в модели наукастинга позволяет получать оценку текущего состояния экономики максимально оперативно.

В работе проводится сравнение следующих моделей: ограниченные и неограниченные MIDAS-модели (модели со смешанными данными), MIDAS-модель с L1-регуляризацией и MFBVAR-модель (байесовская векторная авторегрессия смешанной частоты). Результаты сравниваются с классической авторегрессией для обоснования необходимости использования моделей наукастинга для оперативной оценки макроэкономических показателей. В качестве объясняющих переменных использованы индексы производства по разным отраслям и макропоказатели, характеризующие ВВП России по использованию и его компоненты, а также показатели, полученные на основе анализа тональности новостей, публикуемых рядом крупных СМИ и блогов. В работе также предложен способ оперативной оценки текущего состояния экономики на основе данных только за первый или за первые два месяца рассматриваемого квартала. Использование такого подхода в сочетании с анализом новостного фона позволяет формулировать выводы по текущей экономической ситуации с опережением по сравнению с альтернативными способами наукастинга.

Для формирования переменных, отражающих новостной фон, был использован один из подходов обработки естественного языка (Natural Language Processing), базирующийся на использовании статистической меры TF-IDF. Данная статистическая мера позволяет оценить важность того или иного слова в тексте конкретной новости и соотнести его важность по отношению ко всем новостям. Для расчета данной меры используются 2 основных показателя: частота слова (term frequency, сокр. TF), рассчитываемая как отношение числа вхождений слова к общему числу слов в новости и обратная частота документа (inverse document frequency, сокр. IDF), являющаяся инверсией частоты, с которой слово встречается в наборе новостей. По данной методологии были кластеризованы 500 тысяч новостей и далее рассчитаны следующие показатели: доля новостей с негативным сентиментом (настроением) в общем количестве новостей, доля новостей по теме «Экономика и бизнес» в общем количестве новостей и доля новостей с негативным сентиментом по теме «Экономика и бизнес» в общем количестве новостей по теме «Экономика и бизнес». Показано, что добавление переменных, характеризующих новостной фон, положительно сказывается на точности итоговых моделей.

Результатом исследования становится система наукастинга для ВВП и его компонентов по использованию: для каждой зависимой переменной выбирается лучшая для построения наукаста модель по последним 12 точкам на основе критерия средней абсолютной ошибки (MAE) и среднеквадратичной ошибки прогноза (RMSE).