

УДК 004.89

*Лыченко Н.М., Сорокова А.В., [nlychenko@mail.ru](mailto:nlychenko@mail.ru), [nastusha24sh-g@yandex.com](mailto:nastusha24sh-g@yandex.com)*

*Институт машиноведения и автоматизации НАН КР, Бишкек, Кыргызстан*

## **МОДИФИКАЦИЯ КЛАССИФИКАТОРА ИНДЕКСА КАЧЕСТВА ВОЗДУХА Г. БИШКЕК С УЧЕТОМ ФАКТОРА ЗАГРЯЗНЕНИЯ**

Представлены результаты анализа и использования информации о количестве тонн ежедневно сжигаемого угля на ТЭЦ г. Бишкек для повышения точности прогноза класса индекса качества воздуха (AQI) на основе LSTM-нейронного классификатора. Этот классификатор позволяет в зависимости от метеорологических условий и предшествующей истории значений AQI прогнозировать класс AQI из возможных четырех интегрированных классов:  $AQI \leq 50$  /  $50 < AQI \leq 100$  /  $100 < AQI \leq 150$  /  $AQI > 150$ . Учет фактора загрязнения в качестве дополнительного входа классификатора позволил получить прогноз класса AQI с точностью не менее 80%.

**Ключевые слова:** классификация, прогноз, индекс качества воздуха, LSTM-нейронная сеть, факторы загрязнения

**Введение.** В работах [1, 2] построены модели прогноза концентрации в атмосферном воздухе г. Бишкек твердых частиц размером менее 2.5 микрона (PM 2.5) и прогноза класса индекса качества воздуха (AQI, Air Quality Index) [3] из возможных четырех интегрированных классов:  $AQI \leq 50$  («Хороший»),  $50 < AQI \leq 100$  («Умеренный»),  $100 < AQI \leq 150$  («Нездоровый для чувствительных групп»),  $AQI > 150$  («Нездоровый», «Очень нездоровый», «Опасный»). В качестве признаков (факторов) в этих моделях использована доступная информация [4] о метеорологических параметрах: используются исторические ряды наблюдений за температурой воздуха, температурой точки росы, атмосферным давлением, относительной влажностью воздуха, скоростью ветра, интенсивностью осадков. Эти модели вполне можно использовать для оценки загрязненности воздуха в зависимости от метеофакторов: достигнутая точность прогноза в моделях на основе обобщенно-регрессионной нейронной сети GRNN – около 70%, на основе LSTM-нейронной сети – около 75%. Однако стоит признать, полученные результаты в большой степени зависят от выборки наблюдений и не показывают устойчивости прогноза. Это можно объяснить тем, что использование при построении прогноза только метеофакторов недостаточно, хотя, конечно, синоптические процессы и локальные метеорологические условия во многом определяют поведение поступающих в атмосферу загрязняющих веществ. Вследствие этого построению прогнозных моделей на основе метеофакторов посвящено множество работ, например [5, 6, 7]. Безусловно, важным фактором в процессах загрязнения атмосферного воздуха в городах являются техногенные (антропогенные) источники загрязнения. Поскольку г. Бишкек не располагает в настоящее время крупным промышленным производством, основными факторами загрязнения являются автотранспорт, городская ТЭЦ, котельные и выбросы печей частного сектора в отопительный период. Интересным представляется учесть по возможности эти факторы при построении моделей прогноза.

В апреле 2021 года ОАО «Электрические станции» КР представило Институту машиноведения и автоматизации НАН КР информацию о количестве тонн ежедневно сжигаемого на ТЭЦ г. Бишкек угля и его марке за период с 9 февраля 2019 года по 31 марта 2021 года. В настоящей работе представлены результаты анализа и использования этой информации для модификации модели прогноза класса индекса качества воздуха г. Бишкек на основе LSTM-нейронного классификатора [2].

**2. Данные загрязнения атмосферного воздуха г. Бишкек.** В исследовании использованы данные о загрязнении атмосферного воздуха г. Бишкек, размещенные на сайте [8]. Данные включают индекс качества воздуха AQI и данные о концентрации в воздухе твердых частиц PM2.5 ( $\text{mkg}/\text{m}^3$ ), начиная с 06.02.19 по настоящее время (период измерения – 1 час).

На рисунках 1 и 2 представлены временные ряды (ВР) изменения содержания PM2.5 в атмосферном воздухе и индекса качества воздуха г. Бишкек за период с 09.02.2019 по 31.03.2021 с интервалом в 3 часа (с 30.08.2019 по 04.09.2019 и с 10.12.2020 9.00 по 14.12.2020 12.00 измерения отсутствуют, поэтому на графиках эти измерения представлены средними значениями).

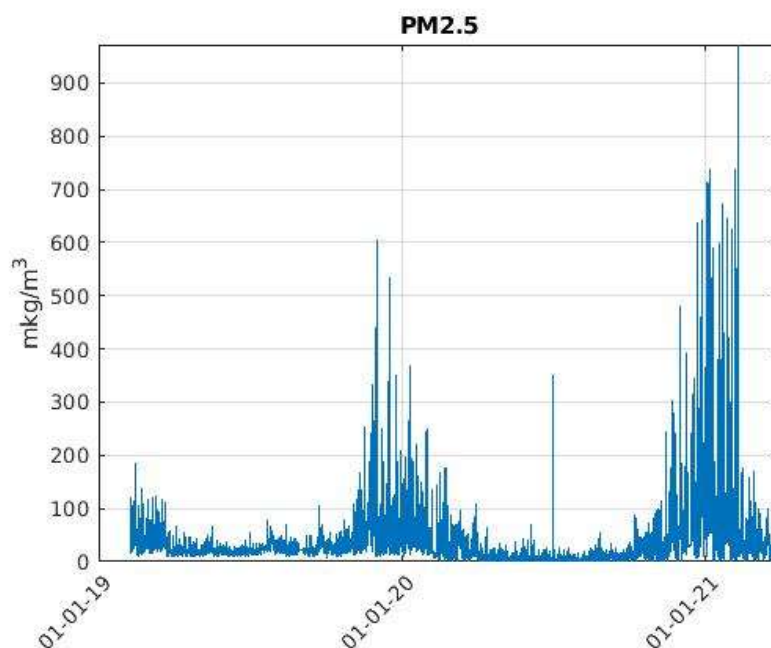


Рисунок 1 – Значения концентраций PM2.5 в воздухе г. Бишкек за период 09.02.2019-31.03.2021

На графиках (рисунки 1,2) вполне отчетливо видна сезонная составляющая: в теплое (летнее) время года средний уровень загрязненности падает, в холодное (зимнее) – повышается. Как показали предыдущие исследования [2], направленные на повышение точности прогноза нейросетевой модели, оправдано связать начало и конец холодного/теплого периодов с началом/ концом отопительных сезонов в городе.

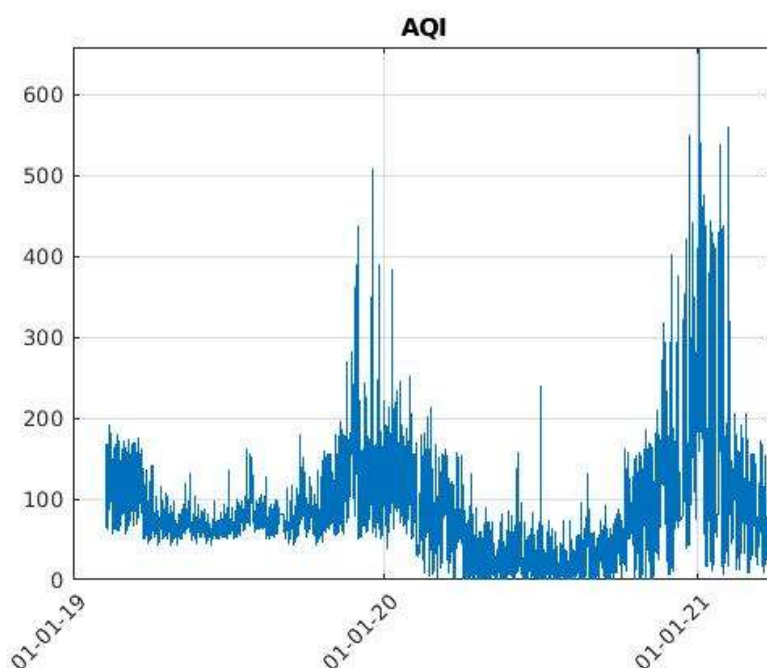


Рисунок 2 – Значения индекса качества воздуха AQI г. Бишкек за период 09.02.2019-31.03.2021

В таблице 1 представлены среднее (математическое ожидание)  $mean$  и стандартное (среднеквадратическое) отклонения  $\sigma$  за весь период наблюдений, а также за периоды 09.02.2019–23.03.2019, 24.03.2019–31.10.2019, 01.11.2019–31.03.2020, 01.04.2020–09.10.2020, 10.10.2020–31.03.2021, соответствующие началу и концу отопительных сезонов в городе (конец отопительного сезона в 2019 году – 22.03.2019, начало – 01.11.2019; конец отопительного сезона в 2020 году – 31.03.2020 (дата приблизительная), начало – 10.10.2020).

На рисунке 3 представлены автокорреляционные функции (АКФ) ВР концентраций PM2.5 и ВР индекса качества воздуха.

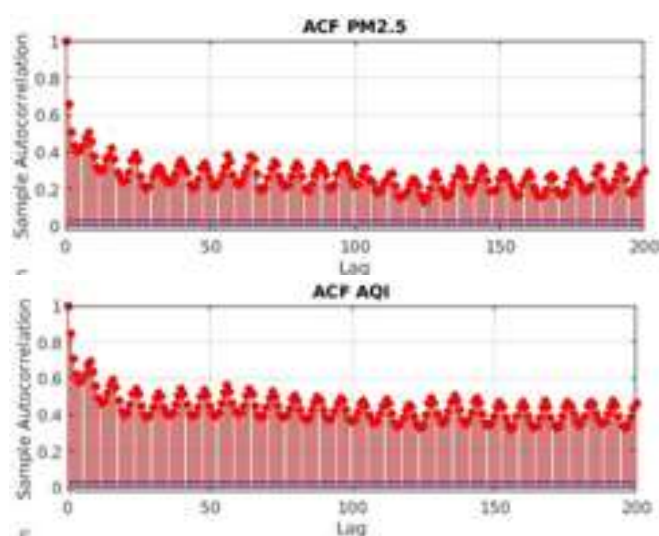


Рисунок 3 – Автокорреляционные функции временных рядов PM2.5 и AQI

Из анализа АКФ следует, что во ВР наблюдений присутствует периодичность в 8 лагов, что соответствует 24 часам. Это объясняется повторением суточной истории в процессах загрязнения: в утренние часы, как правило, уровень загрязнения ниже, чем в вечерние, что в основном объясняется физико-географическими и климатическими условиями города.

Таблица 1 – Средние значения и стандартные отклонения концентраций PM2.5 и AQI

Период наблюдений	PM2.5		AQI	
	mean, mkg/m <sup>3</sup>	$\sigma$ , mkg/m <sup>3</sup>	mean	$\sigma$
09.02.2019-31.03.2021	37.4	65.6	83.9	67.5
09.02.2019-23.03.2019	44.0	23.8	115.1	32.9
24.03.2019-31.10.2019	22.3	10.4	70.0	21.2
01.11.2019-31.03.2020	40.7	122.0	102.2	128.9
01.04.2020-09.10.2020	6.9	8.4	26.0	23.1
10.10.2020-31.03.2021	76.1	115.2	126.9	97.8

**Сравнение летних периодов 2019/20 и 2020/21 годов.** Из графика (рисунок 1) и таблицы 1 видно, что в большинстве случаев значения PM2.5 в летние периоды находятся в пределах принятой в Кыргызстане нормы (35 мкг/м<sup>3</sup>). Однако при более детальном рассмотрении ВР видно, что исключение, как правило, составляет 21 час, когда достаточно много значений эту норму превышают.

Стоит отметить, что введение карантина и ЧС весной 2020 года в связи с распространением в мире COVID-19 привело к значительному (в среднем в три раза) уменьшению концентраций PM2.5, по сравнению с аналогичным (летним) периодом 2019 года. Это связано со снижением производственной и транспортной нагрузки на экологию крупных городов. На рисунке 4 приведены графики кривых, отображающих измеренные значения PM2.5 в г. Бишкек за периоды времени с 1 апреля по 30 сентября 2019 и 2020 гг. (здесь выброс в измерении 01.07.2020, 15.00 (PM2.5=349), в 50 раз превышающий среднее за летний период 2020 г., заменен на среднее значение). Учитывая, что в г. Бишкек практически отсутствуют промышленные предприятия, график концентраций PM2.5 на рисунке 4 можно расценивать как иллюстрацию влияния автотранспорта на уровень загрязнения атмосферного воздуха в городе.

Наличие некоторого количества отрицательных значений PM2.5 за летний период 2020 г. объясняется шумом при измерении, если атмосфера очень чистая (приближается к 0 мкг/м<sup>3</sup>). Поскольку отрицательные числа не имеют физического смысла, при дальнейшей обработке данных они заменены нулями.

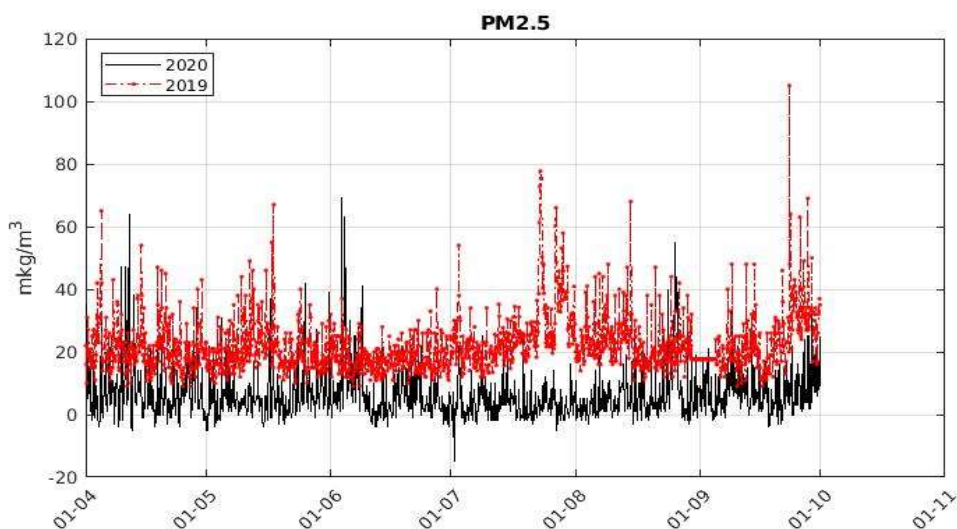


Рисунок 4 – Концентрации PM2.5 в воздухе г. Бишкек в апреле-сентябре 2020 г. и в апреле-сентябре 2019 г.

**Сравнение зимних периодов 2019/20 и 2020/21 годов.** Из графика на рисунке 1 и таблицы 1 видно, что значения PM 2.5 в холодные (зимние) периоды времени

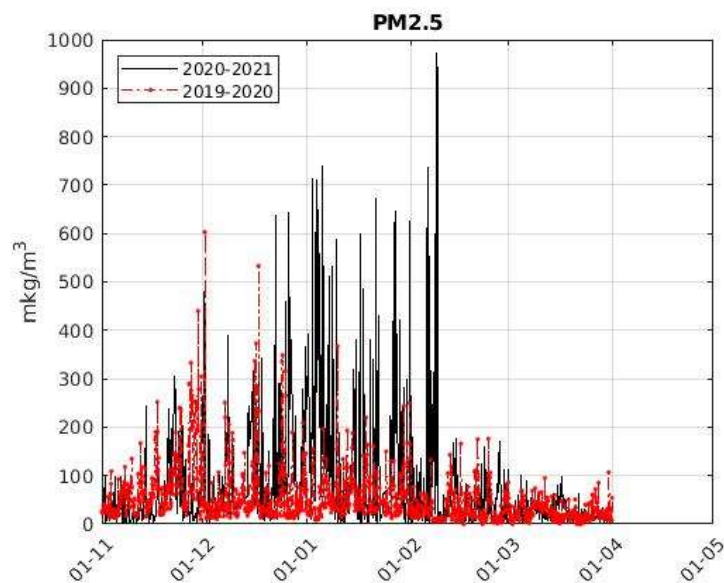


Рисунок 5 – Сравнение зимних периодов 2019/20 гг. и 2020/21 гг.

достаточно резко возрастают и зачастую значительно превышают допустимую норму. Это связано с увеличением загрязняющих атмосферу выбросов городской ТЭЦ и котельных города, производственные мощности которых обусловлены началом и концом отопительных периодов, печей частного сектора, а также с существованием длительных неблагоприятных метеорологических условий, которым свойственны температурные инверсии, слабые ветры, туманы. Стоит отметить (см. таблицу 1), что среднее значение концентрации PM2.5 в зимний период 2020/21 годов в 1.87 раза выше среднего значения концентрации PM2.5 в зимние периоды предыдущих лет. Увеличение загрязненности воздуха

также иллюстрирует рисунок 5, на котором представлены графики PM2.5 зимних периодов 2019/20 и 2020/21 годов.

На первый взгляд это может быть объяснено более низкими в среднем температурами воздуха зимой в 2020/21 гг. (см. таблицу 2).

Таблица 2 – Средняя температура воздуха в зимние месяцы 2020/21 гг.

	2019-2020	2020-2021
Ноябрь	1.4	1.0
Декабрь	1.1	-4.95
Январь	-1.7	-3.8
Февраль	3.7	3.9
Март	8.2	6
Среднее за период	2.48	0.17

Однако проведенный подобно представленному в [9] корреляционный анализ температуры воздуха и концентраций PM2.5 показал, что влияние температуры атмосферного воздуха  $T$  на концентрацию PM2.5 для зимнего периода 2020/21 гг. на треть меньше, чем за предыдущие зимние периоды. В таблице 3 представлены коэффициенты корреляции температуры атмосферного воздуха  $T$ , измеренной в сроки  $i, i-1, i-2, \dots$ , и концентрации PM2.5, измеренной в срок  $i$ . Сроки  $i-1, i-2, \dots$ , соответствуют измерениям на 3, 6 и т.д. часов ранее. Из таблицы видно, что наиболее сильное влияние на концентрацию PM2.5 оказывает температура, измеренная в сроки  $i-4, i-5, i-6$ , при этом коэффициенты корреляции для периода наблюдения 26.11.2020-31.03.2021 меньше, чем для зимних периодов 2019/20 гг. для всех соответствующих сроков измерения. Этот анализ дает основания предположить, что в зимний период 2020/21 гг. появились дополнительные факторы, имеющие отношение к источникам загрязнения.

Таблица 3 – Коэффициенты корреляции  $R(T/PM2.5)$  температуры атмосферного воздуха  $T$ , измеренной в сроки  $i, i-1, i-2, \dots$ , и концентрации PM2.5, измеренной в срок  $i$  для зимних периодов наблюдений

Срок измерения	Период наблюдений	
	09.02.2019 – 23.03.2019 01.11.2019 – 31.03.2020	26.11.2020-31.03.2021
	R (T/ PM2.5)	
i	-0.31	-0.19
i-1	-0.25	-0.15
i-2	-0.29	-0.20
i-3	-0.36	-0.28
i-4	-0.45	-0.35
i-5	-0.49	-0.36
i-6	-0.46	-0.34

**Анализ влияния количества тонн угля, ежедневно потребляемого ТЭЦ г. Бишкек, на среднесуточные значения PM2.5.** Представленные ОАО «Электрические станции» КР данные позволили провести корреляционный и регрессионный анализ среднесуточных концентраций загрязнения воздуха частицами PM2.5 и количества тонн угля, ежедневно потребляемого ТЭЦ г. Бишкек – одного из источников загрязнения атмосферного воздуха города – за период 09.02.2019 – 31.03.2021. Анализ проведен с целью выявления влияния фактора загрязнения на значения концентраций PM2.5 и дальнейшего использования этой информации для построения моделей прогноза, позволяющих предупреждать неблагоприятные экологические ситуации.

В таблице 4 представлена оценка влияния количества тонн угля *Coal*, потребляемого в день *i* на среднесуточную концентрацию *PM2.5*, измеренную в день *i*, а также показано влияние инерционного фактора, т.е. влияние количества тонн угля *Coal*, потребляемого в дни *i-1*, *i-2* (соответственно одним и двумя днями ранее), на среднесуточную концентрацию *PM2.5*, измеренную в день *i*. Из таблицы видно, что влияние практически безынерционное – коэффициенты корреляции для различных дней практически не изменяются, а также стоит отметить, что влияние достаточно существенное – коэффициент корреляции  $R=0.56$ .

Таблица 4 – Коэффициенты корреляции *R* количества тонн угля, потребляемого ТЭЦ в день *i*, *i-1*, *i-2*, на среднесуточную концентрацию *PM2.5* в *i*-й день.

День измерения	R (Coal/ PM2.5)
<i>i</i>	0.556
<i>i-1</i>	0.550
<i>i-2</i>	0.546

Имея наборы данных о среднесуточных концентрациях загрязнения воздуха и количестве тонн угля, ежедневно потребляемого ТЭЦ, можно так же, как и для метеорологических данных [9], рассчитать статистические зависимости между переменными, используя метод регрессионного анализа [10]. На рисунке 6 представлен график среднесуточного изменения *PM2.5* в зависимости от количества тонн угля, потребляемого ТЭЦ г. Бишкек в те же сутки, соответствующий полученному уравнению линейной регрессии:

$$PM2.5(i) = 0.764 + 0.0111 * Coal(i).$$

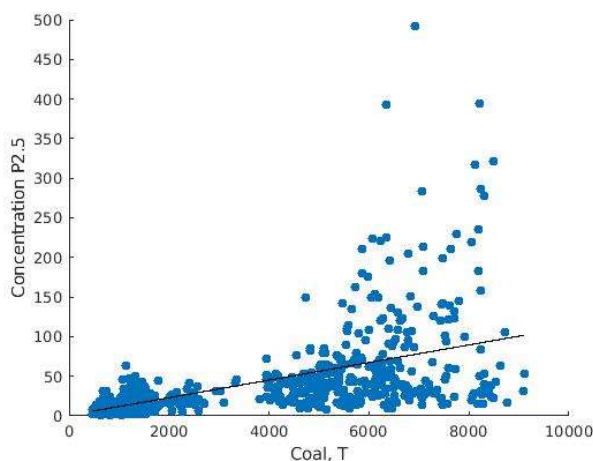


Рисунок 6 – Зависимость среднесуточных концентраций *PM2.5* от количества тонн угля, ежедневно потребляемого ТЭЦ:  $PM2.5 = 0.0111 * Coal + 0.76$  для периода 09.02.2019 – 31.03.2021.

Таким образом, установленная корреляция может служить дополнительной исходной информацией для формирования более эффективных моделей прогнозирования загрязнения воздуха на основе искусственных нейронных сетей [11, 12], в частности, LSTM-нейросетевого классификатора.

**Модификация LSTM-нейросетевого классификатора.** Структура классификатора аналогична структуре, выбранной в [2]. Только теперь к факторам, определяющим вход-

ной вектор классификатора (который в [2] включал температуру воздуха, температуру точки росы, атмосферное давление на уровне станции, относительную влажность, скорость ветра, интенсивность осадков), прибавился еще один фактор – количество тонн угля, сжигаемого ежедневно на городской ТЭЦ. Стоит отметить, что этот фактор также несет информацию о сезонности, поскольку границы сезонов связаны с отопительным периодом; поэтому в новом классификаторе фактора-признака сезонности нет.

Выходной вектор классификатора, как и в [2], определяется параметрами, которые характеризуют вероятности отнесения выхода классификатора к тем или иным классам индекса качества воздуха AQI из возможных четырех интегрированных классов:  $AQI \leq 50$  («Хороший»),  $50 < AQI \leq 100$  («Умеренный»),  $100 < AQI \leq 150$  («Нездоровый для чувствительных групп»),  $AQI > 150$  («Нездоровый», «Очень нездоровый», «Опасный»).

В исследовании возможностей сети по прогнозированию класса AQI варьировались следующие параметры:

S – длина последовательности векторов исторических данных входных векторов классификатора;

P – глубина прогноза (на сколько шагов вперед прогнозируется AQI). Шаг прогноза – 3 часа.

Программная реализация инструмента для проведения экспериментов извлекает данные, нормализует данные по заданной функции нормализации и по заданным S и P генерирует все возможные последовательности входных векторов классификатора длиной S и сопоставляет им значения AQI в моменты времени, удаленные на P шагов от момента последних значений этих последовательностей. В таблице 5 представлены результаты прогноза с учетом данных о количестве тонн угля, ежедневно сжигаемого на городской ТЭЦ.

Таблица 5 – F-мера [2] точности модели прогнозирования четырех классов AQI в зависимости от длины последовательности входных векторов (S) и глубины прогноза (P) для выборки наблюдений 09.02.2019-31.03.2020+01.08.2020-31.03.2021 с учетом и без учета данных ТЭЦ.

S	P	F-мера 4 класса (без ТЭЦ)	F-мера 4 класса (ТЭЦ)
8	4	0.7641	0.7945
8	8	0.7449	0.7707
8	16	0.7286	0.7749
8	24	0.7270	0.7567
8	32	0.7139	0.7736
12	4	0.7669	0.8023
12	8	0.7728	0.8013
12	16	0.7571	0.7930
12	24	0.7556	0.7876
12	32	0.7456	0.8010

Как видно из таблицы, в зависимости от глубины прогноза P и длины выборки наблюдений S, подаваемых на вход нейросетевого классификатора, точность прогноза класса AQI с учетом данных ТЭЦ выше на 3–5%, нежели без учета этих данных.

**Заключение.** Учет нового фактора – информации о количестве тонн сжигаемого на ТЭЦ г. Бишкек угля – позволил повысить точность прогнозирования класса AQI из четырех интегрированных классов ( $AQI \leq 50$  /  $50 < AQI \leq 100$  /  $100 < AQI \leq 150$  /  $AQI > 150$ ) до 80%. Это дает основание утверждать, что учет этого фактора так же, как и метеофакторов, имеет



большое значение для решения важной задачи прогноза уровня загрязнения воздуха в городе. Адекватный прогноз AQI позволит правительственным организациям в случае очень низкого качества воздуха, когда может быть нанесен значительный вред здоровью граждан, объявлять о чрезвычайной экологической ситуации и принимать меры по ограничению выбросов основных источников загрязнения воздуха до тех пор, пока не уменьшатся опасные метеорологические условия.

**Благодарности.** Авторы выражают благодарность отделу по связям с общественностью ОАО «Электрические станции» КР за представленную информацию о количестве тонн ежедневно сжигаемого на ТЭЦ г. Бишкек угля и его марке за период с 9 февраля 2019 года по 31 марта 2021 года.

### **Литература**

1. Великанова Л.И., Лыченко Н.М. Мультирегрессионные и обобщенно-регрессионные нейросетевые модели краткосрочного прогноза загрязнения PM<sub>2.5</sub> в г. Бишкек с учетом метеорологических параметров// Проблемы автоматизации и управления. – N2. – Бишкек: Илим, 2019. – С. 42–51.
2. Лыченко Н.М., Сороковая А.В. Прогнозирование классов индекса качества воздуха г. Бишкек с учетом новых данных 2020-2021 гг. на базе LSTM-нейросетевого классификатора// Проблемы автоматизации и управления. – 2021. – № 2 (41). – С. 89 – 99.
3. Air Quality Index (AQI) - A Guide to Air Quality and Your Health. US EPA. 9 December 2011.
4. Сайт «Расписание погоды rp5.ru» Архив погоды в Бишкеке [https://rp5.ru/%D0%90%D1%80%D1%85%D0%B8%D0%B2\\_%D0%BF%D0%BE%D0%B3%D0%BE%D0%B4%D1%8B\\_%D0%B2\\_%D0%91%D0%B8%D1%88%D0%BA%D0%B5%D0%BA%D0%B5](https://rp5.ru/%D0%90%D1%80%D1%85%D0%B8%D0%B2_%D0%BF%D0%BE%D0%B3%D0%BE%D0%B4%D1%8B_%D0%B2_%D0%91%D0%B8%D1%88%D0%BA%D0%B5%D0%BA%D0%B5) (дата обращения: 30.07.2021)
5. Burhan Baran. Prediction of Air Quality Index by Extreme Learning Machines// 2019 International Artificial Intelligence and Data Processing Symposium (IDAP) DOI: 10.1109/IDAP.2019.8875910
6. Soubhik Mahanta, T. Ramakrishnudu; Rajat Raj Jha; Niraj Tailor. Urban Air Quality Prediction Using Regression Analysis//TENCON 2019 - 2019 IEEE Region 10 Conference (TENCON). Date of Conference: 17-20 Oct. 2019 DOI: 10.1109/TENCON.2019.8929517
7. M. Sharma; S. Aggarwal; P. Bose; A. Deshpande. Meteorology-based forecasting of air quality index using neural network//IEEE International Conference on Industrial Informatics, 2003. INDIN 2003. Proceedings. DOI: 10.1109/INDIN.2003.1300360
8. AirNow Department of State // [https://airnow.gov/index.cfm?action=airnow.global\\_summary#U.S.\\_Department\\_of\\_State\\_Bishkek](https://airnow.gov/index.cfm?action=airnow.global_summary#U.S._Department_of_State_Bishkek), (дата обращения: 30.07.2021).
9. Лыченко Н.М. Регрессионный анализ метеорологических факторов и концентраций частиц PM<sub>2.5</sub> в атмосферном воздухе г. Бишкек// Проблемы автоматизации и управления. – N2. – Бишкек: Илим, 2019. – С. 5–15.
10. Бокс Д., Дженкинс Т. Анализ временных рядов: прогноз и управление. – М.: Мир, 1974. – 242 с.

11. X. Zhao, R. Zhang, J.-L. Wu, P.-C. Chang. A Deep Recurrent Neural Network for Air Quality Classification // *Journal of Information Hiding ultimedia Signal Processing*. – V.9, N.2, March 2018.
12. Understanding LSTM Networks. – URL: <http://colah.github.io/posts/2015-08-Understanding-LSTMs/> (дата обращения 14.09.2021).