



**К.К. Бесерра Муриель**

*Байкальский государственный университет,  
г. Иркутск, Российская Федерация*

## **Прогнозирование будущей стоимости колумбийского инвестиционного фонда с помощью рекуррентных нейронных сетей LSTM (LSTM)**

**Аннотация.** Рекуррентные нейронные используются для прогнозирования цен в финансовых временных рядах. Одной из широко используемых моделей рекуррентных нейронных сетей является модель LSTM (Long Short-Term Memory), предложенная Сеппом Хохрейтером и Юргеном Шмидхубером в их работе LONG SHORT-TERM MEMORY, опубликованной в 1997 г. Эта модель решает проблему долговременной памяти рекуррентных нейронных сетей путем добавления ячейки селективной памяти, которая действует как «фильтр», выбирая, какую информацию важно сохранить, а какая информация не имеет значения и может быть отброшена. Цель данной работы — проанализировать, насколько оригинальная LSTM-модель пригодна для прогнозирования цены колумбийского инвестиционного фонда без учета экзогенных переменных (экономических новостей, макро- и микроэкономических переменных, базовых активов, таких как акции, опционы, облигации и т.д.), влияющих на его поведение.

В данном исследовании использовались три фонда с преобладающим составом акций на колумбийском рынке, которые имеют более высокий уровень волатильности, чем государственные или частные облигации.

**Ключевые слова.** Искусственный интеллект, глубокое обучение, инвестиционные фонды, долгосрочная память, LSTM, взаимные фонды, рекуррентные нейронные сети, RNN.

**Информация о статье.** Дата поступления: 27 февраля 2024 г.; дата принятия к публикации: 12 марта 2024 г.; дата онлайн-размещения: 30 марта 2024 г.

Original article

**C.C. Becerra Muriel**  
*Baikal State University,  
Irkutsk, Russian Federation*

## **Forecasting the Future Value of a Colombian Investment Fund with LSTM Recurrent Neural Networks (LSTM)**

**Abstract.** Recurrent neural networks are a tool that is currently used in time series, a widespread use of these networks is the forecasting of future prices in financial time series. One widely used recurrent neural network model is the LSTM (Long Short-Term Memory) model, proposed by Sepp Hochreiter and Jürgen Schmidhuber in their paper called LONG SHORT-TERM MEMORY published in 1997. This model solves the long term memory problem of recurrent neural networks by adding a selective memory cell which acts as a "filter" to choose what kind of information is important to keep and what kind of information is irrelevant and can be discarded.

This paper seeks to analyze whether the original LSTM model is relevant to forecast the price of a Colombian mutual fund without taking into account exogenous variables (economic news, macroeconomic and microeconomic variables, underlying assets such as stocks, options, bonds, etc.) that affect its behavior.

In this study, three funds with a majority composition of shares in the Colombian market were taken, which have higher volatility levels than public or private bonds.

**Keywords.** Artificial Intelligence, Deep Learning, Investment Funds, Long Short-Term Memory, LSTM, Mutual Funds, Recurrent Neural Networks, RNN.

**Article info.** Received 27 February, 2024; Accepted 12 March, 2024; Available online 30 March, 2024.

## Введение

Паевые инвестиционные фонды в Колумбии, называемые финансовым регулятором "Superintendencia Financiera de Colombia" "fondos de inversión en Colombia", представляют собой инвестиционную систему, объединяющую различные финансовые инструменты, такие как акции, облигации, товары, ETF, и другие инструменты, такие как корпоративные долги, корпоративные счета, кредиты сотрудникам, выданные в качестве гарантии оплаты и т.д., аналогичные инструментам взаимных фондов, торгуемых на международных рынках<sup>1</sup>.

В настоящее время "Фонды коллективных инвестиций" являются одним из самых недооцененных финансовых инструментов во всем финансовом секторе Колумбии, несмотря на то, что являются инвестиционными инструментами, которые предлагают более высокие ставки доходности, чем те, которые могут предложить сберегательные счета или срочные депозитные сертификаты CDT, превышающие инфляцию в стране [1]. Это представлено в основном по нескольким причинам, среди которых выделяется расчет годовой доходности, вызывающий путаницу у международных инвесторов из-за расчета, произведенного в Колумбии, который отличается от международных стандартов. Ниже показано, как рассчитывается доходность<sup>2</sup>.

$$Rp = \left( \frac{VUO_y}{VUO_x} \right)^{\left( \frac{365}{n} \right)} - 1, \quad (1)$$

где  $Rp$  — эффективная годовая доходность за период между днями  $x$  и  $y$ ,  $VUO_y$  — значение NAV за день  $y$ ,  $VUO_x$  — значение NAV для дня  $x$ ,  $n$  — количество дней, прошедших между  $x$  и  $y$ .

<sup>1</sup> Decreto 1242 de 2013 por el cual se sustituye la Parte 3 del Decreto número 2555 de 2010 en lo relacionado con la administración y gestión de los Fondos de Inversión Colectiva. URL: <https://www.funcionpublica.gov.co/eva/gestornormativo/norma.php?i=89110>.

<sup>2</sup> Circular Básica Contable y Financiera (Circular Externa 100 de 1995). Capítulo XI. Valoración de fondos de inversión colectiva y de las inversiones que conforman sus portafolios. URL: <https://www.superfinanciera.gov.co/publicaciones/15466/normativanormativa-generalcircular-basica-contable-y-financiera-circular-externa-de-15466>.

По сути, эта формула рассчитывает вариацию между точкой  $x$  и точкой  $y$ , затем формула приводит годовой коэффициент к 365 календарным дням, а затем вычитает единицу, чтобы получить результат, который интерпретируется как индикатор более высокой доходности, чем та, которая имела место в действительности.

Однако, местные инвесторы, обладающие избыточной ликвидностью, воздерживаются от вложений в такие инструменты из-за чрезмерного недоверия, существующего на колумбийском фондовом рынке, учитывая многочисленные случаи мошенничества и кражи средств у инвесторов управляющими взаимными фондами, при этом ни один государственный регулирующий орган не смог предпринять никаких эффективных действий, которые могли бы обеспечить безопасность инвесторов [2].

Это произошло из-за отсутствия в Колумбии метода надзора и объективного мониторинга поведения инвестиционных фондов, который бы успокоил инвесторов и позволил им принять превентивные меры на рынке ценных бумаг в случае какой-либо аномалии и избежать корректирующих действий, которые мало чем помогут колумбийскому рынку и усилят недоверие среди инвесторов и потенциальных инвесторов, рассматривающих эти инструменты как нежизнеспособную альтернативу для своих вложений.

Поэтому в Колумбии многие независимые компании создали независимые методы мониторинга, позволяющие следить за курсом и поведением колумбийских инвестиционных фондов. Американский сайт [carterascolectivas.com](http://carterascolectivas.com) проводит параллельную оценку эффективности фондов на основе методологии, широко используемой в Соединенных Штатах.

Искусственный интеллект не является инструментом, который официально используется в колумбийском регуляторе "Superintendencia Financiera de Colombia", поэтому в данной статье мы предлагаем использовать модель рекуррентной нейронной сети под названием LSTM (Long Short-Term Memory).

Модель рекуррентной нейронной сети LSTM была предложена Сеппом Хохрайтером и Юргеном Шмидхубером в их работе под названием LONG SHORT-TERM MEMORY [3], опубликованной в 1997 г. Эта модель предлагает решение проблемы памяти для обычных рекуррентных нейронных сетей, которые имеют или работают только с кратковременной памятью. В данном случае используется концепция ячейки состояний (забывание, обновление, выход), которая позволяет сортировать, обновлять и удалять информацию в соответствии с ее первоначальной целью.

В каждом нейроне добавлена ячейка памяти, эта память имеет вес  $w = 1$ , что позволяет закрепить память сети в долгосрочной перспективе. В дополнение к входным элементам рекуррентной нейронной сети также имеется дополнительный

вход, который является ячейкой памяти, прохождение информации в эту ячейку контролируется воротами и вычислительными блоками (см. рис. 1). Ниже подробно объясняется, как работают эти ворота.

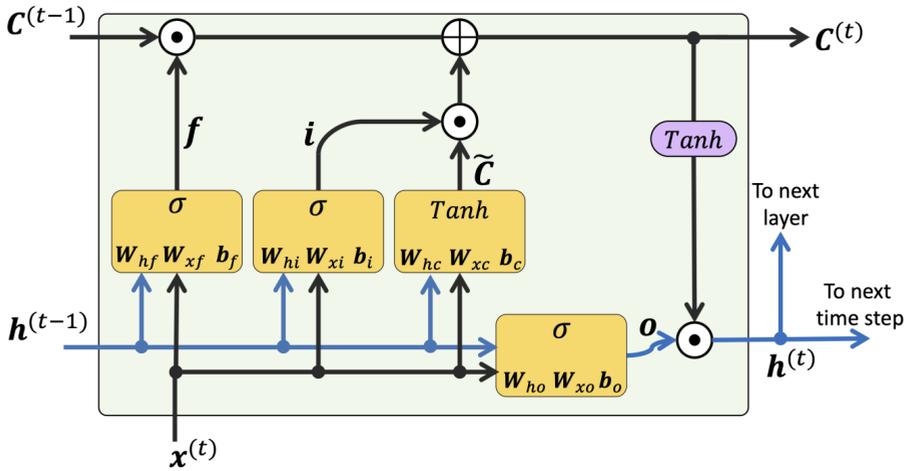


Рис. 1. Рекуррентная нейронная сеть [4]

Ворота забывания: эти ворота принимают решение о том, какая информация может быть обработана, а какая должна быть удалена, имея некоторые заранее установленные критерии, расчет производится по следующей формуле

$$f_t = \sigma(W_{xf}x^t + W_{hf}x^{t-1} + b_f). \quad (2)$$

Эти ворота изначально не входили в модель Сеппа Хохрайтера и Юргена Шмидхубера, а были предложены в более поздней модели под названием *Learning to forget: Continual Prediction with LSTM* [4], предложенной Феликсом А. Герсом, Юргеном Шмидхубером и Фредом Камминсом в 1999 г.

Входные ворота: эти ворота позволяют вводить новую информацию и обновлять состояние ячейки памяти. Они разрабатываются с учетом следующих формул:

$$i_t = \sigma(W_{xi}x^t + W_{hi}x^{t-1} + b_i); \quad (3)$$

$$g_t = \tanh(W_{xg}x^t + W_{hg}x^{t-1} + b_g). \quad (4)$$

Выходные ворота: Эти ворота решают, как обновлять значения внутри скрытых блоков, их расчет производится по следующей формуле

$$\sigma_t = \sigma(W_{x\sigma}x^t + W_{h\sigma}x^{t-1} + b_\sigma). \quad (5)$$

**Цель работы.** Определить с помощью модели глубокого обучения LSTM мониторинг цены NAV взаимных фондов в Колумбии с учетом правил, установленных колумбийскими регулирующими органами для расчета цены.

### Описание данных колумбийских инвестиционных фондов для анализа в модели LSTM

Колумбийские взаимные фонды оцениваются по методологии NAV (Net Asset Value), которая учитывает стоимость AUM (Asset Under Management) фонда, умноженную на количество паев в обращении.

Любой фонд, который начинает свою деятельность, должен начать со стоимости пая или NAV (Net Asset Value) в COP (колумбийское песо) \$10.000, на закрытие первого дня NAV такого инвестиционного фонда должна составлять COP \$10.000, после чего необходимо учитывать следующие формулы [5]:

1. Расчет стоимости фонда AUM

$$VFC_{t-1} + RD_t + VENP_t, \quad (6)$$

где  $VFC_{t-1}$  — частичное закрытие фонда коллективных инвестиций за день  $t$ ,  $RD_t$  — результаты за день  $t$  (Доходы — Расходы),  $VENP_t$  — стоимость выпуска новых паев в день  $t$  (применимо только для фондов, учрежденных в соответствии с колумбийским законодательством);

2. Расчет стоимости пая фонда NAV

$$VUO = \frac{PCF_t}{NUC_{t-1} + NENP_t}, \quad (7)$$

где  $VUO_t$  — стоимость единицы для операций в день  $t$ ,  $PCF_t$  — частичное закрытие фонда коллективных инвестиций за день  $t$ ,  $NUC_{t-1}$  — количество паев паевого инвестиционного фонда на момент закрытия торгов в предыдущий день ( $t-1$ ),  $NENP_t$  — количество новых акций.

Для этого анализа использовались следующие фонды:

1. Инвестиционный фонд Accicolf Vanguardia Ordinary Shares:

Описание профиля фонда: *Портфель Фонда коллективных инвестиций состоит из обыкновенных акций, привилегированных акций и/или облигаций с обязательной конвертацией, которые могут быть или не быть зарегистрированы в Национальном реестре ценных бумаг и эмитентов или включены в листинг иностранной системы торговли ценными бумагами*<sup>3</sup>.

2. Инвестиционный фонд Ishares Msci Colcap Equity Fund:

<sup>3</sup> Fondo de Inversión Colectiva Cerrada Accicolf Vanguardia Ordinary Shares. URL: <https://www.fiduoccidente.com/empresas/corporativo-e-institucional/accicolf>.

Описание профиля фонда: *iShares MSCI COLCAP ETF стремится достичь инвестиционных результатов, соответствующих доходности индекса MSCI COLCAP*<sup>4</sup>.

### 3. Инвестиционный фонд Btg Pactual Acciones Colombia:

Описание фонда: *Алгоритм реализован следующим образом: Цель открытого паевого инвестиционного фонда BTG Pactual Acciones Colombia — предоставить своим подписчикам долгосрочный инструмент для инвестиций в колумбийский рынок акций*<sup>5</sup>.

#### **Реализация модели LSTM:**

1. Набор данных настраивается с использованием стоимости пая фонда (NAV), которая является ежедневной величиной, изменяющейся ежедневно только один раз, поэтому у нее нет таких показателей, как цены High, Low, Open, а есть только цена Close<sup>6</sup>.

2. Данные разделены на две части (83 % обучения и 17 % проверки для фондов ISHARES и PACTUAL, 80 % обучения и 20 % проверки для фонда VANGUARDIA).

3. В целях обучения данные нормализуются до диапазона от 0 до 1, после чего возвращаются в исходное состояние.

4. Обучение данных происходит путем взятия блоков по 50 данных, чтобы предсказать данные 51. Этот блок из 50 данных хранится в переменной X, а данные 51 будут храниться в переменной Y, которые, в свою очередь, будут выходом LSTM-сети.

5. Наборы, созданные в списках, масштабируются таким образом, чтобы в сети был известен вектор 50 x 1.

6. Контейнер создается с помощью модуля Sequential библиотеки Keras, выходной (результатирующий) слой — с помощью модуля Dense.

7. Методом проб и ошибок было определено число в 50 нейронов.

8. Модель составляется с использованием оптимизатора rmsprop и определением функции ошибки mse (средняя квадратичная ошибка).

9. Для обучения модели используется 50 реплик (эпох), а также партии из 64 данных.

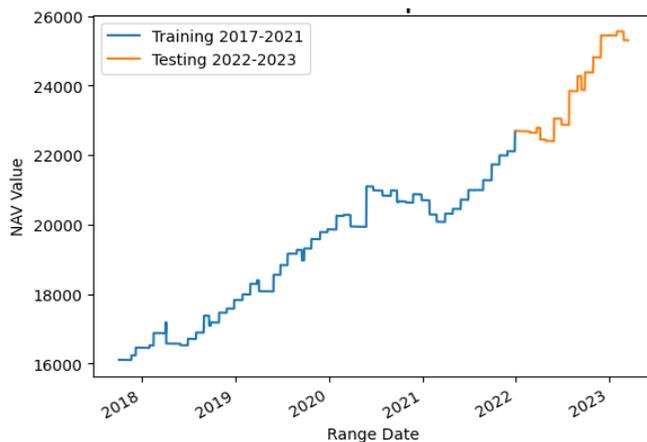
10. После обучения модель проверяется, поэтому данные, которые были подготовлены для обучения, имели состав 50 данных на блок, поэтому проверочный набор также будет иметь такой же состав, это делается с помощью функции предсказания.

<sup>4</sup> Fondo Bursátil iShares MSCI COLCAP. URL: <https://www.blackrock.com/co/productos/251708/ishares-fondo-burstil-ishares-colcap-fund>

<sup>5</sup> Fondo de inversion Colectiva Abierto Btg Pactual Acciones Colombia. URL: <https://www.btgpactual.com.co/es/que-hacemos/productos/fondos-de-inversion-colectiva-fics/renta-variable/fic-acciones-colombia>

<sup>6</sup> Rentabilidades de los Fondos de Inversión Colectiva (FIC). Datos Abiertos. URL: [https://www.datos.gov.co/Hacienda-y-Credito-P-blico/Rentabilidades-de-los-Fondos-de-Inversi-n-Colectiv/qhpu-8ixx/data\\_preview](https://www.datos.gov.co/Hacienda-y-Credito-P-blico/Rentabilidades-de-los-Fondos-de-Inversi-n-Colectiv/qhpu-8ixx/data_preview)

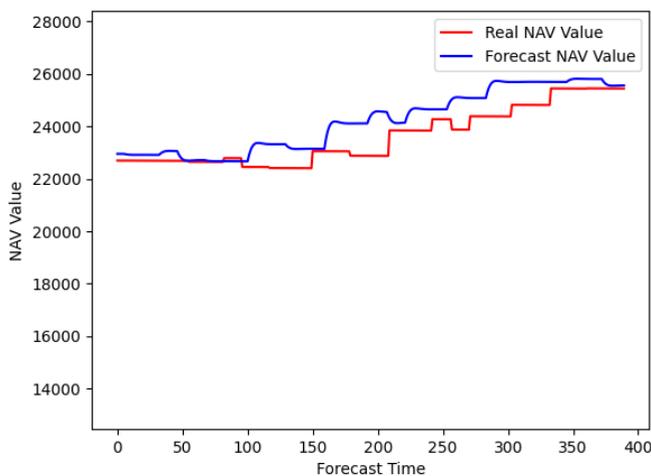
Ниже приведены графики трех анализируемых фондов:  
 1. *Инвестиционный фонд Accicolf Vanguardia Ordinary Shares*  
 График динамики NAV с 2018 по 2023 г. приведены на рис. 2.



**Рис. 2. График динамики NAV с 2018 по 2023 г.**

Источник: URL: <https://colab.research.google.com/drive/1L-kAdLaejUhf697-yIf-bOdO6dP9JgPPo?usp=sharing>.

График, показывающий сравнение фактических данных NAV и данных NAV, предсказанных моделью, приведен на рис. 3.  
 Средняя абсолютная ошибка составляет 479,2055194794872  
 Средняя квадратичная ошибка равна 393132.9325392043  
 Средняя еквадратичная ошибка равна 627.0031359883333



**Рис. 3. График сравнения фактического значения NAV и прогнозируемого значения NAV**

2. Инвестиционный фонд акций Ishares Msci Colcap

График динамики NAV с 2018 по 2023 г. приведен на рис. 4.

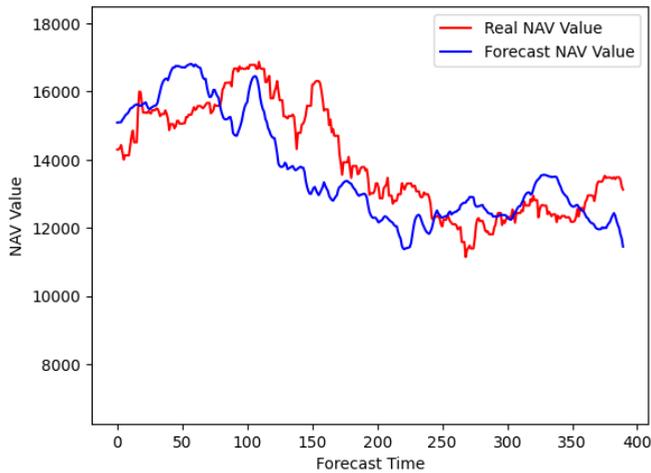


Рис. 4. График динамики NAV с 2018 по 2023 г.

Источник: URL: <https://colab.research.google.com/drive/1aUPwvhq8S8EqpJBRy1czwxD6di2m42RK?usp=sharing>.

График, показывающий сравнение фактических данных NAV и данных NAV, предсказанных моделью, приведен на рис. 5. Средняя абсолютная ошибка составляет 959,7953683205128  
Средняя квадратичная ошибка равна 1368290.7305573323  
Средняя еквадратичная ошибка равна 1169.7395994653393

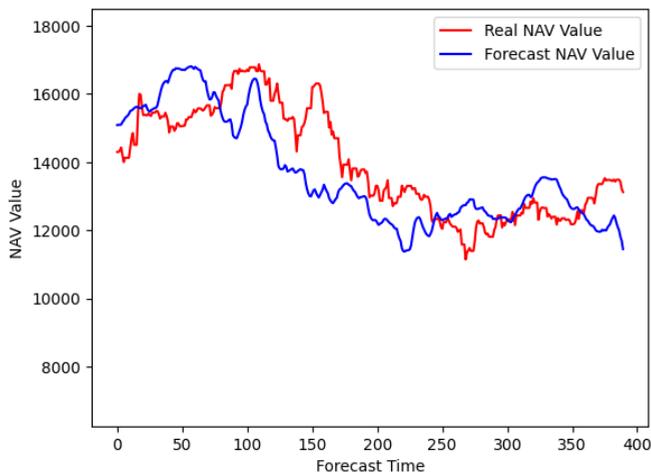


Рис.5. График сравнения фактического значения NAV и прогнозируемого значения NAV

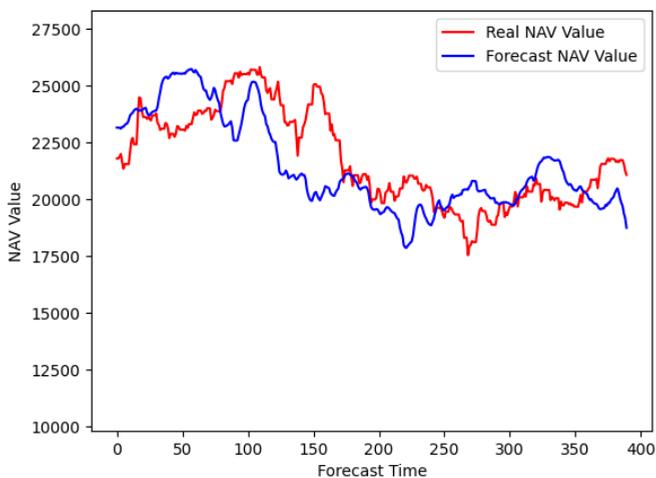
### 3. Инвестиционный фонд Btg Pactual Acciones Colombia



**Рис. 6. График динамики NAV с 2018 по 2023 г.**

Источник: URL: [https://colab.research.google.com/drive/1gBx4pYutVQqEWGr0EkVjxAzKkPmb\\_hFw?usp=sharing](https://colab.research.google.com/drive/1gBx4pYutVQqEWGr0EkVjxAzKkPmb_hFw?usp=sharing).

График динамики NAV с 2018 по 2023 г. приведен на рис. 6. График, показывающий сравнение фактических данных NAV и данных NAV, предсказанных моделью, приведен на рис. 7. Средняя абсолютная ошибка составляет 1425.7098085057694. Средняя квадратичная ошибка равна 3083635.0470006275. Средняя квадратичная ошибка равна 1756.028202222455.



**Рис. 7. График сравнения фактического значения NAV и прогнозируемого значения NAV**

Параметры модели, приведены на рис. 8.

Layer (type)	Output Shape	Param #
lstm (LSTM)	(None, 50)	10400
dense (Dense)	(None, 1)	51

=====  
 Total params: 10451 (40.82 KB)  
 Trainable params: 10451 (40.82 KB)  
 Non-trainable params: 0 (0.00 Byte)

**Рис. 8. Параметры модели**

### Выводы

Предложенный метод можно применять для интервальной оценки. По результатам измерений ошибок в модели LSTM можно определить, что она может быть полезна, если используется для создания диапазона допустимых цен.

Анализ секторов экономики, частью которых являются базовые активы фондов, необходимо добавить в качестве экзогенных переменных в ранее созданную LSTM-модель.

Объем выборки фондов, анализируемых в данной модели, составляет в среднем 2 550 данных. Учитывая характер анализируемых фондов, в которых нет больше информации, чем та, что была получена для данного анализа, необходимо рассмотреть дополнительные алгоритмы, которые могут дополнить модель LSTM для получения лучших результатов.

Для получения надежных результатов при работе с таким типом данных модель LSTM необходимо настроить.

### References

1. Ramírez-Córdoba G. *La inversión colectiva en Colombia: caracterización y análisis*. 2014. 17 p. Available at: <https://repository.eia.edu.co/handle/11190/702>.
2. Tito-Añamuro J.A. Corrupción privada: un estudio de la ausencia de reglas de derecho privado, desde el caso Interbolsa. *Vniversitas*, 2015, vol. 64, no. 131, pp. 440–466. DOI: <http://dx.doi.org/10.11144/Javeriana.vj131.cpea>.
3. Sepp Hochreiter, Jürgen Schmidhuber. Long Short-Term Memory. *Neural Computation*, 1997, vol. 9, iss. 8. Available at: <https://doi.org/10.1162/neco.1997.9.8.1735>.
4. Gers F.A. Learning to Forget: Continual Prediction with LSTM. *9<sup>th</sup> International Conference on Artificial Neural Networks: ICANN*. 1999. DOI: <https://doi.org/10.1049/cp:19991218>.
5. Raschka S., Mirjalili V. *Python Machine Learning*. 2<sup>nd</sup> ed. Mumbai, Packt Publishing Ltd, 2017. 270 p.

### Информация об авторе

**Бесерра Муриель Кристиан Камило** — аспирант, кафедры математических методов и цифровых технологий, Байкальский государственный университет, г. Иркутск, Российская Федерация, e-mail: bcristiancamilo@mail.ru.

### Information about the Author

**Cristian C. Becerra Muriel** — PhD Student, Department of Mathematical Methods and Digital Technologies, Baikal State University, Irkutsk, Russian Federation, e-mail: bcristiancamilo@mail.ru.

### Для цитирования

Бесерра Муриель К.К. Прогнозирование будущей стоимости колумбийского инвестиционного фонда с помощью рекуррентных нейронных сетей LSTM (LSTM) / К.К. Бесерра Муриель. — DOI 10.17150/2713-1734.2024.6(1).78-88. — EDN LTAZEM // *System Analysis & Mathematical Modeling*. — 2024. — Т. 6, № 1. — С. 78–88.

### For Citation

Becerra Muriel C.C. Forecasting the Future Value of a Colombian Investment Fund with LSTM Recurrent Neural Networks (LSTM). *System Analysis & Mathematical Modeling*, 2024, vol. 6, no. 1, pp. 78–88. (In Russian). EDN: LTAZEM. DOI: 10.17150/2713-1734.2024.6(1).78-88.