УДК 519.6

# ПРИМЕНЕНИЕ АЛГОРИТМА МАШИННОГО ОБУЧЕНИЯ ДЛЯ РЕАЛИЗАЦИИ ПРОЕКТОВ ПРИ ПОДГОТОВКЕ ВЕБ-РАЗРАБОТЧИКОВ

**В. А. Кочурко[1)], Д. В. Филимонов[2)], С. В. Харланов[3)], И. Р. Лукьянович[4)]**

[1), 2), 3), 4)] *Белорусский государственный университет, Беларусь, Минск,*
[1)] *vladislav.kochurko@gmail.com,* [2)] *dzfilimonau@gmail.com,*
[3)] *s.v.harlanov@gmail.com,* [4)]*lukianinna12345@gmail.com*

Рассмотрено применение алгоритма контролируемого обучения, который использован при решения задачи регрессии в разработке приложения для мониторинга и предсказания ситуации COVID-19 – статистических данных о заболевших, умерших, выздоровевших, вакцинированных – на ближайшую неделю. Спроектированы, разработаны и интегрированы в приложение база данных, модель предсказания данных, back-end и front-end составляющие приложения. Применение алгоритмов машинного обучения в курсовом и дипломном проектировании имеет большое прикладное значение как для создания веб-приложений, направленных на прогнозирование и поддержку принятия решений, так и для повышения профессиональных компетенций веб-разработчиков математического профиля.

***Ключевые слова:*** алгоритмы машинного обучения; метод случайного леса; нереляционная база данных MongoDB; Scikit-learn; Pandas; библиотека математических алгоритмов NumPy; математические специальности.

# APPLYING A MACHINE LEARNING ALGORITHM FOR THE IMPLEMENTATION OF PROJECTS IN THE PREPARATION OF WEB DEVELOPERS

**V. A. Kachurka[1)], D. V. Filimonov[2)], S. V. Charlanau[3)], I. R. Lukyanovich[4)]**

[1), 2), 3), 4)] *Belarussian state university, Belarus, Minsk,*
[1)] *vladislav.kochurko@gmail.com,* [2)] *dzfilimonau@gmail.com,*
[3)] *s.v.harlanov@gmail.com,* [4)]*lukianinna12345@gmail.com*

The application of the supervised learning algorithm, which was used to solve the regression problem in the development of an application for monitoring and predicting the situation of COVID-19 - statistical data on sick, dead, recovered, vaccinated - for the next week, is considered. A database, a data prediction model, back-end and front-end components of the application were designed, developed and integrated into the application. The use of machine learning algorithms in course and diploma design is of great practical

importance for creating web applications aimed at forecasting and decision support and to improve the professional competencies of web developers with a mathematical profile.

***Keywords:*** machine learning algorithms, random forest method, MongoDB non-relational database, Scikit-learn, Pandas, NumPy math algorithms library, math majors.

## Введение

Пандемия COVID-19 стала причиной серьезных социально-экономических последствий, включая крупнейшую мировую рецессию. Дезинформация о вирусе распространилась через социальные сети и средства массовой информации. Это сопровождалось паникой, усугубляя и без того очень неблагоприятную ситуацию в мире. Для предоставления реальных данных о COVID-19 были разработаны специальные приложения, где каждый человек мог посмотреть данные по любой стране в мире. Приходится, однако, констатировать, что количество веб-приложений, позволяющих прогнозировать развитие эпидемиологической ситуации, является недостаточным. Цель данной работы – разработать приложение для мониторинга и предсказания ситуации COVID-19 на ближайшую неделю с помощью нейронной сети.

Актуальность практического применения математических методов в курсовых и дипломных веб-проектах состоит в повышении качества обучения специалистов математического профиля, как с точки зрения прикладных аспектов использования изучаемого математического аппарата, так и для освоения и разработки методов компьютерного моделирования.

Используемые инструменты разработки и созданное веб-приложение позволят создавать подобные системы для отслеживания и прогнозирования распространения массовых заболеваний и других явлений, для описания которых подходит модель случайного леса.

## Основные задачи реализации проекта

Для решения поставленной цели необходимо было решить следующие задачи: найти источники достоверных данных об эпидемической ситуации в различных странах [1], [2]; определить технологии, необходимые для реализации проекта, спроектировать и реализовать базу данных; разработать back-end составляющую приложения; разработать модель предсказания данных и интегрировать ее в back-end составляющую; разработать front-end составляющую приложения; провести тестирование и оптимизацию веб-приложения.

Анализ веб-приложения [1] позволил определить основные составляющие интерфейса создаваемого ресурса: интерактивную карту, панель

с общей информацией и списком стран, страницу с информацией о стране (рис. 1).

*Рис. 1.* Главная страница по завершении загрузки

В качестве основных показателей были выделены: новые смерти, число заболевших, смертей и выздоровевших за весь период пандемии, население страны, число людей, болеющих в настоящее время, уровень смертности, число заболевших на миллион, число смертей на миллион, а также число вакцинированных людей. визуализация данных о COVID-19 в виде графиков (рис. 2), (рис. 3)

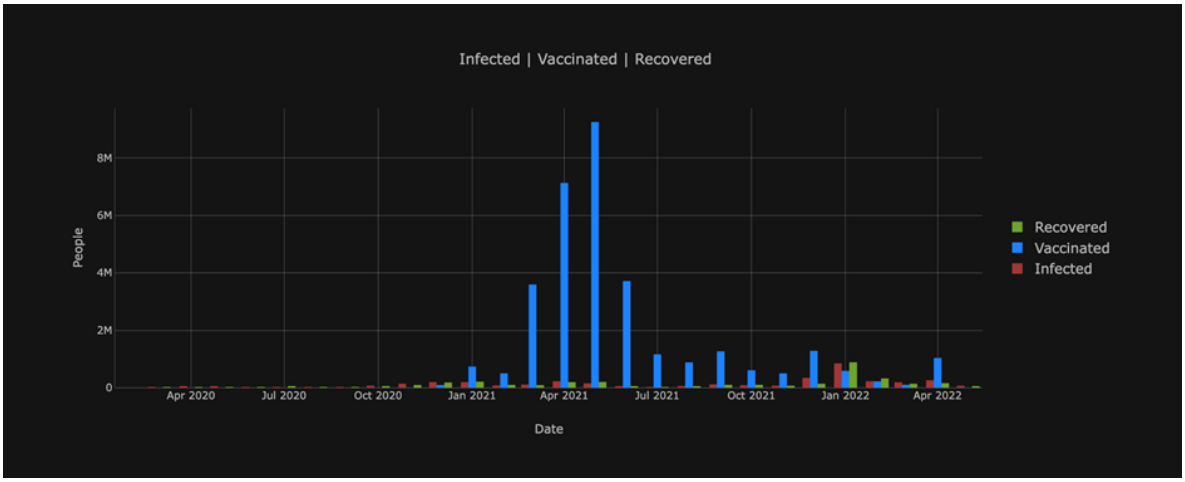


*Рис. 2.* Линейный график предсказания.

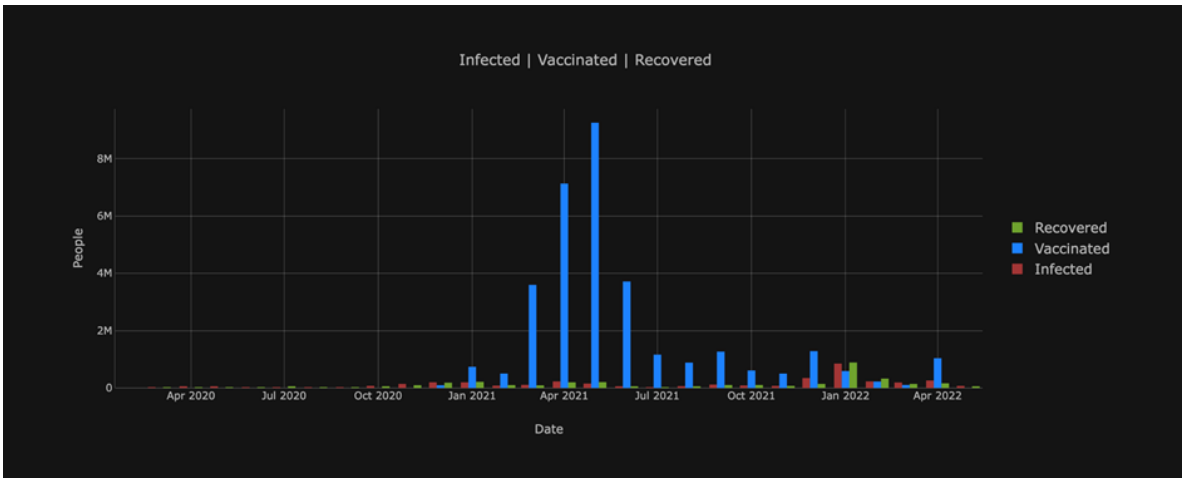


*Рис. 3.* Гистограмма “Infected | Vaccinated | Recovered”

Графики имеют собственную функциональность. При наведении на график появляется всплывающая подсказка с датой и значением параметра на выбранную дату. Пользователь может масштабировать графики. С помощью интерактивной легенды можно включать/отключать отображение некоторых параметров. В верхнем правом углу находится меню, с помощью него пользователь может вернуться в исходное состояние и сохранить график в виде картинки в формате .png.

**Важнейшие технологии реализации проекта**

Для разработки front-end части приложения применен фреймворк Angular и с языком программирования TypeScript [3]. Построение диаграмм и графиков обеспечило использование библиотек D3.js, Plotly.js [4]. Интерактивная карта реализована на основе библиотеки Leaflet [5].

Для разработки Back-end составляющей приложения была использована серверная платформа NodeJS и ее фреймворк Express [6], а также TypeScript. База данных проекта реализована с помощью нереляционной СУБД MongoDB.

Сборка проекта выполнена на основе Webpack [7]. Анализ кода и исправление ошибок выполнено с помощью ESLint [8] с конфигурацией airbnb-typescript/base.

Нейронная сеть для предсказания эпидемиологической ситуации реализована на языке программирования Python с применением библиотеки машинного обучения Scikit-learn [9].

В качестве вспомогательных библиотек для работы с данными и их анализа во время построения модели используются Python-библиотеки Pandas и NumPy, ориентированные на Data Science и Machine Learning.

## Выбор алгоритма реализации модели предсказания

Из постановки задачи следует что алгоритм реализации модели – контролируемое обучение. Это значит, что обучение алгоритма контролируется разработчиком в процессе работы. Для этого маркируются данные и устанавливаются границы, которым должен следовать алгоритм.

Учитывая, что результатом работы алгоритма является численное выражение одной из опций на основе других, можно сделать вывод, что поставленная задача – задача регрессии. В задачах регрессии программа машинного обучения должна оценить и понять взаимосвязи между переменными. Предмет регрессионного анализа – одна зависимая переменная и набор других изменяющихся переменных. Это делает анализ особенно полезным для прогнозирования и предсказаний.

Для выбора подходящего алгоритма существенно, что у нас есть несколько входных параметров, характер зависимости которых непостоянен и не всегда определяется. Исходя из этого можно выделить, что будет удобно использовать алгоритм, эффективно работающий с несколькими входными параметрами, оценивающий каждый из параметров, но не позволяющий доминировать определенному параметру, и который получает эффективное решение при большом количестве возможных предсказаний.

На основании вышесказанного выбран алгоритм случайного леса. Данный алгоритм является весьма распространенным в машинном обучении, его эффективность напрямую зависит от количества решающих деревьев в ансамбле [10].

## Заключение

Несмотря на то, что эпидемия COVID-19 практически сошла на нет, ее протекание и последствия свидетельствуют о важности разработки моделей прогнозирования и веб-приложений на их основе.

## Библиографические ссылки

1. Covid-19 API . [Электронный ресурс] – Режим доступа: https://coronavirus.app – Дата доступа: 13.04.2023.

2. Country API . [Электронный ресурс] – Режим доступа: https://restcountries.com/v3.1 – Дата доступа: 13.04.2023.

3. Документация к языку программирования TypeScript. [Электронный ресурс] – Режим доступа: https://www.typescriptlang.org/docs/ – Дата доступа: 13.04.2023.

4. Документация D3.js . [Электронный ресурс] – Режим доступа: https://d3js.org/ – Дата доступа: 13.04.2023.

5. Документация Leaflet. [Электронный ресурс] – Режим доступа: https://leafletjs.com/ – Дата доступа: 13.04.2023.
6. Документация Express . [Электронный ресурс] – Режим доступа: https://expressjs.com/ru/ – Дата доступа: 13.04.2023.
7. Документация Webpack. [Электронный ресурс] – Режим доступа: https://webpack.js.org/ – Дата доступа: 13.04.2023
8. Документация ESlint. [Электронный ресурс] – Режим доступа: https://eslint.org/ – Дата доступа: 13.04.2023.
9. Документация Scikit-learn. [Электронный ресурс] – Режим доступа: https://scikit-learn.org/stable/ – Дата доступа: 13.04.2023.
10. Алгоритм случайного леса. [Электронный ресурс] – Режим доступа: https://proglib.io/p/mashinnoe-obuchenie-dlya-nachinayushchih-algoritm-sluchaynogo-lesa-random-forest-2021-08-12 - Дата доступа: 13.04.2023