АНАЛИЗ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В КОНТЕКСТЕ ПРОГНОЗИРОВАНИЯ ИНСУЛЬТОВ

Виктория ТРЕБИШ,

студентка II год, Факультет Математики и Информатики

CZU: 004.89:616.831-005 trebisvictoria@outlook.com

По данным Всемирной организации здравоохранения, инсульт – вторая по значимости причина смерти и третья по значимости причина инвалидности в мире. 6,5 миллионов человек ежегодно умирают от инсульта, а в настоящее время в мире проживает более 101 миллиона человек, перенёсших инсульт. Расходы, связанные с инсультом только в США составили почти 56,5 млрд долларов в период с 2018 по 2019 год. Предупреждения возникновения инсульта является одной из основных задач современной медицины. Исходя из этого, основной целью данного исследования является разработка модели машинного обучения, способной с высокой точностью на ранних стадиях прогнозировать вероятность появления инсульта, с целью оперативного принятия необходимых медицинских мер.

Основные этиологические факторы, предрасполагающие к развитию инсульта и рассматриваемые в контексте всех его типов, включают: возраст, пол, курение, сахарный диабет, индекс массы тела, заболевания сердца, стресс. Эти факторы и были включены в процесс отбора признаков при разработке модели, с дополнительным включением нескольких других: семейное положение, трудовой статус и тип места жительства.

Эффективность использования искусственного интеллекта при оценке риска развития сердечно-сосудистых заболеваний была обоснована в работах [2-4].

1. Метод *k*-ближайших соседей (KNN)

KNN классифицирует новые образцы, опираясь на классы *k*-ближайших элементов из обучающего набора данных, относя объект к классу, представленному наибольшим числом соседей. С целью улучшения результатов классификации вводится масштабирование данных и взвешивание соседей, так чтобы от их приближенности зависел их уровень вклада в предсказание.

2. Дерево решений

Решающие деревья представляют собой невероятно мощное семейство моделей в машинном обучении, по причине их способности прослеживать нелинейные зависимости. По своей сути эти модели – это предикторы h: $x \rightarrow y$, которые предсказывают метку, ассоциированную собразцом x, проходя от корня дерева к узлу с помощью применения последовательности простых решающих правил. Этот процесс в некотором смысле согласуется с естественным для человека процессом принятия решений.

3. Случайный лес

Один из способов уменьшения опасности переобучения деревьев заключается в построении случайного леса, одного из популярнейших способов применить

«*бэггинг*». Идея состоит в том, что мы имеем *N* решающих деревьев и тогда решением нашей задачи классификации будет результат голосования. Для лучшей обобщающей необходимо достичь некоррелированности моделей через использование бутстрапирования и метода случайных подпространств.

По причине самой высокой эффективности этот метод был выбран за основной.

Таблица 1. Точность алгоритмов

Алгоритм машинного обучения	**Точность (%)**
KNN	94
Решающее дерево	92
Случайный лес	96

В данной работе был проведен анализ существующих методов прогнозирования инсульта, предложены различные методы машинного обучения и выбран наиболее оптимальный метод.

Книги и монографические издания:

1. ВОРОНИНА, В. В., МИХЕЕВ, А. В., ЯРУШКИНА, Н. Г., СВЯТОВ, К. В. Теория и практика машинного обучения: учебное пособие. – Ульяновск: УлГТУ, 2017. – С. 9-67.
2. ШАЛЕВ-ШВАРЦ Ш., Бен-ДАВИД Ш. Идеи машинного обучения: от теории к алгоритмам / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2019. – С. 176-257.
3. ЭЛБОН К. Машинное обучение с использованием Python. Сборник рецептов: Пер. с англ. — СПб.: БХВ-Петербург, 2019. – С. 73-104.
4. КЭВИН МЭРФИ П. Вероятностное машинное обучение: введение / пер. с англ. А. А. Слинки- на. – М.: ДМК Пресс, 2022. – С. 225-246.
5. ОРЛАНД П. Математические алгоритмы для программистов. 3D-графика, машинное обучение и моделирование на Python. — СПб.: Питер, 2023. – С. 589-658.

Recomandat

***Gheorghe CĂPĂȚÂNĂ**, dr. hab., prof. univ.*