|  |  |
| --- | --- |
| **Gerb-BMSTU_01** | **Министерство науки и высшего образования Российской Федерации**  **Федеральное государственное бюджетное образовательное учреждение**  **высшего образования**  **«Московский государственный технический университет**  **имени Н.Э. Баумана**  **(национальный исследовательский университет)»**  **(МГТУ им. Н.Э. Баумана)** |

Физкультурно-оздоровительный факультет

Кафедра «Здоровьесберегающие технологии и  
 адаптивная физическая культура»

**РЕФЕРАТ**

**«*Анализ факторов, влияющих на потребление алкоголя школьниками старших классов, методом вариации коэффициентов функции активации нейронной сети Хопфилда*»**

Выполнил студент: \_\_***Белоусов Евгений Александрович***\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*фамилия, имя, отчество*

Группа: \_\_\_\_***ИУ5-61Б***\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Проверил доцент кафедры АФК, к.психол.н**\_\_\_\_\_\_\_\_\_\_\_\_Худышева М.К.**

*подпись, дата*

Оценка \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ Дата \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

*2020 г.*

# Оглавление

[Оглавление 1](#_Toc39279542)

[Введение 2](#_Toc39279543)

[Основная часть 3](#_Toc39279544)

[1. Существующие подходы к решению задачи 4](#_Toc39279545)

[2. Метод вариации порога функции активации нейронной сети Хопфилда 5](#_Toc39279546)

[3. Другие исследования 7](#_Toc39279547)

[Вывод 9](#_Toc39279548)

[Список литературы 10](#_Toc39279549)

[Приложение 11](#_Toc39279550)

# 

# Введение

Сложно представить современный мир без алгоритмов машинного обучения [1]. Они постепенно проникают во все сферы жизни, и находят применение в самых неожиданных областях. Эти алгоритмы обрабатывают огромные объемы данных, выявляя в них почти невидимые человеческому глазу зависимости.

Среди алгоритмов машинного обучения большое распространение получили нейронные сети [2]. Они весьма эффективно справляются с большим классом задач, недоступных другим алгоритмам.

Однако, понять, чему именно научилась нейронная сеть, какие связи выстроились между нейронами, на какие атрибуты сеть обращает внимание в первую очередь - сложная и почти невыполнимая задача. Эта задача может встать, например, при проектировании системы обработки информации в реальном времени, когда получение того или иного атрибута связано с определенными денежными или временными затратами. Или, например, при разработке программ по борьбе с инсультом, где определение наиболее критических факторов поможет направить большие точечные усилия на их устранение.

**Цель работы:** исследовать факторы, влияющие на употребление алкоголя старшеклассниками.

**Задачи работы:**

1) рассмотреть существующие подходы к решению задачи;

2) рассмотреть метод вариации коэффициентов функции активации нейронной сети Хопфилда;

3) сравнить полученные результаты с другими исследованиями в этой области.

# Основная часть

В качестве выборки будем анализировать датасет “Student alcohol consumption” [3], предоставленный сайтом kaggle. В нем находятся данные, полученные в ходе анкетирования учащихся курсов математики и португальского языка в средней школе. Параметры, которыми описываются школьники в датасете приведены в приложении.

Факторы, представленные в датасете “Student alcohol consumption”, приведены в приложении. Из них нас не интересует название школы, сразу избавимся от него.

Также в датасете употребление алкоголя в выходные дни и в рабочие представлено в виде двух отдельных порядковых признаков. Чтобы упростить задачу заменим их одним бинарным признаком alc:

И в дальнейшем будем анализировать именно его.

## Существующие подходы к решению задачи

Поставленную задачу можно рассматривать как задачу отбора информативных признаков. Существует несколько способов ее решения. В рамках этой работы рассмотрим два из них: с помощью регуляризации Лассо и с помощью решающего дерева.

При построении линейных моделей может применяться регуляризация Лассо, ее особенность – отсеивание неинформативных признаков. Соответственно, если постепенно увеличивать коэффициент регуляризации, можно пронаблюдать, какие признаки окажутся более важными для модели, а какие – менее.

Недостатком этого метода является необходимость построения линейной модели, которая может описывать только линейные зависимости в данных. Для анализа нашего датасета данный способ не подходит.

Другим, более распространенным методом является построение ансамбля алгоритмов на основе решающих деревьев. Данный метод может работать с любыми типами данных. Чем чаще признак встречается в полученных деревьях решений, тем он более важный.

Так, например, этот способ считает, что употребление алкоголя старшеклассниками больше всего зависит от частоты гуляния с друзьями и от возраста. Данные утверждения выглядят вполне логичными.

Стоит отметить, что существуют и другие методы, но в рамках данной работы они рассматриваться не будут.

## Метод вариации порога функции активации нейронной сети Хопфилда

Еще одним способом отбора информативных признаков является метод вариации коэффициентов функции активации нейронной сети Хопфилда. Для его применения, мы должны обучить нейронную сеть Хопфилда [4] этим данным. Возможно, для этого придется преобразовать данные к виду, пригодному для использования с сетью Хопфилда.

Обучив сеть, мы запускаем алгоритм вариации порога функции активации нейронной сети Хопфилда, получая на его выходе список атрибутов, используемый сетью при вынесении решения. При этом список является отсортированным в порядке убывания или возрастания важности атрибута.

Анализируя данный список, мы можем сделать вывод о важности того или иного фактора.

Данный метод предложен в работе [5], для описания процессов работы мозга. В этой работе мы постарались найти более прикладное применение для данного метода.

Для применения алгоритма вариации порога функции активации нейронной сети Хопфилда необходимо:

1. Обучить нейронную сеть Хопфилда.
2. Сильно завысить порог функции активации. Для этого надо посчитать максимальное значение, получаемое при перемножении матрицы весов W нейронной сети Хопфилда на вектор, заполненный единицами и поставить это значение в качестве порога.
3. Подать на распознавание сети вектор, состоящий из -1, произвести распознавание.
4. Получить некий образ.
5. Немного понизить порог функции активации.
6. Повторять действия 3-5 пока порог функции активации не станет равным изначальному с отрицательным знаком.
7. Анализируя список, полученный образов, можно заметить, что вначале мы получаем образы, состоящие из всех -1, затем появляется одна 1 (самый важный атрибут), с понижением порога появится еще одна единичка (второй по важности атрибут) и так далее. В конце мы получим образ, состоящий из всех 1.

Таким образом, мы получим атрибуты характерные для объектов нашей выборки.

Итак, поделим полную выборку студентов на тех, которые употребляют алкоголь (alc = 1), и тех, которые его не употребляют (alc = 0), и применим к обоим группам метод вариации коэффициентов функции активации нейронной сети Хопфилда, реализованный с помощью кода, описанного в [6].

В результате, для обоих групп самым важным признаком является желание получить высшее образование. Однако, для старшеклассников, не употребляющих алкоголь вторым по важности признаком является отсутствие проваленных экзаменов, затем, посещение детского сада в детстве, отсутствие романтических отношений и более трех человек в семье. В то время, как для употребляющих алкоголь характерно наличие доступа к интернету из дома, совместное проживание родителей и то, что на дорогу из дома до школы, они тратят меньше часа.

Интересно, что одними из наименее характерных факторов для употребляющих алкоголь старшеклассников являются посещение детского сада и отсутствие романтических отношений (то есть по сути их присутствие). А для не употребляющих – затраты на дорогу до школы меньше часа.

## Другие исследования

Выбранная нами тема является популярной. Существует много научных работ в данной области. Большинство из них основывается на результатах статистических исследований и корреляционных анализов.

Так, например, в работе [7] авторы разделяют все факторы, влияющие на формирование потребности употребления алкоголя школьниками/учащимися на 5 групп:

1. медико-биологические факторы;
2. факторы, характеризующие семью и взаимоотношения в ней;
3. факторы, характеризующие социально-гигиенические и финансовое обеспечение;
4. признаки, характеризующие социальное окружение и поведение окружающих;
5. факторы диванного поведения школьников.

Результаты исследования показывают, что наибольшее число достоверно влияющих факторов приходится на группы семейных факторов, а также социального окружения и поведения.

Авторами утверждается, что уже в 13-летнем возрасте школьники ищут место психологической комфортности и зачастую находят такую возможность в окружении друзей, в местах отдыха и часто в неблагополучной социальной среде.

В научной работе [8] авторы также анализируют факторы, влияющие на потребление алкоголя. Они утверждают, что индивиды с более высоким уровнем образования менее склонны к потреблению алкоголя, так как они лучше прогнозируют свое будущее (однако в России доля пьющих среди высокообразованных выше, но они пьют в меньших объемах, а также другие напитки).

Также авторы опровергают мнение, что в семьях с одним родителем дети будут пить больше, хотя отмечают, что дети в только что разведенных семьях склонны к большему употреблению алкоголя. Другим фактором, имеющим влияние на потребление алкоголя среди молодежи, называют наличие постоянного партнера. Еще один фактор – тесные отношения с ровесниками. Авторы находят неудивительным и тот факт, что на потребление алкоголя влияют гендер, возраст и место жительства.

К сожалению, в рассмотренных работах не упоминается ни слова про время, которое учащиеся тратят на дорогу от школы до дома, посещение ими детского сада и наличие доступа к Интернету из дома.

Вероятно, что, так как мы рассматриваем в работе относительно свежие исследования, ко времени их написания Интернет перестал быть чем-то необычным и уже появился у всех. В связи с чем, возможно, этот пункт не учитывался авторами в исследованиях, так как перестал быть информативным.

Посещение детского сада также скорее всего просто не вошло в исследование.

Время, которое учащийся тратит на дорогу до школы, скорее всего связано с другими факторами. Чем меньше времени школьник тратит на дорогу до школы, тем больше свободного времени у него остается, которое можно потратить, например, на общение с друзьями.

Так как в анализе использовалась относительно небольшая выборка, не стоит забывать о возможном шуме в датасете, который мог случайно поднять важность какого-либо признака.

# Вывод

В данной работе было проведено исследование факторов, влияющих на употребление алкоголя старшеклассниками.

Для этого было использовано два метода машинного обучения: построение дерева принятия решений с последующей оценкой важности его параметров и метод вариации коэффициентов функции активации нейронной сети Хопфилда.

Был предложен и подробно описан метод вариации порога функции активации нейронной сети Хопфилда.

Полученные в ходе исследования результаты не противоречат здравому смыслу и подтверждаются другими исследователями в данной области, поэтому их вполне можно считать достоверными.

# Список литературы

1. Жерон, Орельен Прикладное машинное обучение с помощью Sickit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем. 2018. 688с
2. *Norbert Wiener.* Cybernetics: Or Control and Communication in the Animal and the Machine. — 2nd revised ed.. — Paris: Hermann & Cie, Camb. Mass. (MIT Press), 1961. — [ISBN 978-0-262-73009-9](https://ru.wikipedia.org/wiki/Служебная:Источники_книг/9780262730099). Первое издание — 1948
3. Fabio Pagnotta, Hossain Mohammad Amran. Student Alcohol Consumption <https://www.kaggle.com/uciml/student-alcohol-consumption> (дата обращения: 29.04.2020).
4. Hopfield J.J. Learning algorithms and probability distributions in feedforward and feed-back networks. PNAS December 1, 1987 84 (23), pp.8429–8433. https://doi.org/10.1073/pnas.84.23.8429
5. Введенский В.Л., Ежов А.А. Ритмы мозга и самовоспроизведение информации. Природа. 1990. №4. С. 33–44.
6. Гапанюк Ю.Е., Белоусов Е.А., Попов И.А. Реализация нейронной сети Хопфилда на языке PYTHON 3. XVIII Всероссийская научная конференция «Нейрокомпьютеры и их применение». Тезисы докладов. – М: ФГБОУ ВО МГППУ, 2020. С. 44-45.
7. Почитаева И.П., Люцко В.В., Почитаева И.П., Люцко В.В. ОСНОВНЫЕ ФАКТОРЫ, ВЛИЯЮЩИЕ НА ПОТРЕБЛЕНИЕ АЛКОГОЛЯ/НАРКОТИКОВ ШКОЛЬНИКАМИ/УЧАЩИМИСЯ И СТУДЕНТАМИ КОСТРОМСКОЙ ОБЛАСТИ // Современные проблемы науки и образования. – 2015. – № 5.;  
   URL: http://www.science-education.ru/ru/article/view?id=21898 (дата обращения: 29.04.2020).
8. Артемьева Н., Богданов М., Лебедев Д., Назимова А. Факторы, определяющие тип потребления алкоголя среди студентов. Вестник общественного мнения. 2017. С 203-232.

# Приложение

# Факторы, описанные в датасете «Student Alcohol Consumption»

1. school - student's school (binary: 'GP' - Gabriel Pereira or 'MS' - Mousinho da Silveira)
2. sex - student's sex (binary: 'F' - female or 'M' - male)
3. age - student's age (numeric: from 15 to 22)
4. address - student's home address type (binary: 'U' - urban or 'R' - rural)
5. famsize - family size (binary: 'LE3' - less or equal to 3 or 'GT3' - greater than 3)
6. Pstatus - parent's cohabitation status (binary: 'T' - living together or 'A' - apart)
7. Medu - mother's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
8. Fedu - father's education (numeric: 0 - none, 1 - primary education (4th grade), 2 – 5th to 9th grade, 3 – secondary education or 4 – higher education)
9. Mjob - mother's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
10. Fjob - father's job (nominal: 'teacher', 'health' care related, civil 'services' (e.g. administrative or police), 'at\_home' or 'other')
11. reason - reason to choose this school (nominal: close to 'home', school 'reputation', 'course' preference or 'other')
12. guardian - student's guardian (nominal: 'mother', 'father' or 'other')
13. traveltime - home to school travel time (numeric: 1 - 1 hour)
14. studytime - weekly study time (numeric: 1 - 10 hours)
15. failures - number of past class failures (numeric: n if 1<=n<3, else 4)
16. schoolsup - extra educational support (binary: yes or no)
17. famsup - family educational support (binary: yes or no)
18. paid - extra paid classes within the course subject (Math or Portuguese) (binary: yes or no)
19. activities - extra-curricular activities (binary: yes or no)
20. nursery - attended nursery school (binary: yes or no)
21. higher - wants to take higher education (binary: yes or no)
22. internet - Internet access at home (binary: yes or no)
23. romantic - with a romantic relationship (binary: yes or no)
24. famrel - quality of family relationships (numeric: from 1 - very bad to 5 - excellent)
25. freetime - free time after school (numeric: from 1 - very low to 5 - very high)
26. goout - going out with friends (numeric: from 1 - very low to 5 - very high)
27. Dalc - workday alcohol consumption (numeric: from 1 - very low to 5 - very high)
28. Walc - weekend alcohol consumption (numeric: from 1 - very low to 5 - very high)
29. health - current health status (numeric: from 1 - very bad to 5 - very good)
30. absences - number of school absences (numeric: from 0 to 93)