

Санкт-Петербургский Государственный университет
Факультет прикладной математики и процессов управления

Проектная работа по курсу «Неклассические логики»
«Прогнозирование успеваемости студента средствами нечеткой логики»

Работу выполнил
студент группы 19Б11
Адамович Анатолий Олегович

Санкт-Петербург
2022

Содержание

1. Введение	2
2. Постановка задачи и методы её решения.....	3
3. Математическое обоснование	5
3.1. Традиционный подход и нечеткая логика.....	5
3.2. Архитектура нечеткой системы	5
3.3. Лингвистические переменные	6
3.4. База правил	9
3.5. Работа и качество модели	9
3.6. Проблемы и возможные улучшения	11
4. Заключение	13
5. Список литературы.....	14

1. Введение

Довольно часто при поступлении в университет абитуриенты сталкиваются с различными проблемами (трудная адаптация к новым условиям жизни, фрустрация, разочарование в чем-либо), следствием которых в некоторых случаях являются ещё более серьёзные проблемы: перегорание, потеря интереса к учебе, обесценивание престижа высшего образования в целом.

Чтобы оказывать влияние на учебный процесс, необходимо заранее предотвращать проблемы, описанные выше. Для этого необходимо иметь объективную оценку текущей ситуации: следить за настроением студентов, их потребностями и трудностями. На основе этих данных можно определить студентов, находящихся на грани срыва (то есть тех, чья успеваемость либо уже низкая, либо скоро упадет), и персонально работать с ними, чтобы сохранить их мотивацию и помочь им успешно справиться с учебной нагрузкой [1].

Целью данной работы является создание модели прогнозирования, которая на основе связи между успеваемостью студентов и психологическими факторами даст возможность предвидеть проблемы, влияющие на качество получаемого образования и репутацию ВУЗа, чтобы в дальнейшем предпринять меры, направленные на повышение управляемости образовательного процесса.

2. Постановка задачи и методы её решения

Часто можно заметить, что при одних и тех же внешних условиях учебной деятельности студенты проявляют себя по-разному: одни охотно работают над овладением знаний, в то время как другие значительно снижают свою активность [2]. Для объяснения этого феномена стоит принять во внимание тот факт, что однообразность внешних условий мало говорит о психологической стороне вопроса, а именно в ней и состоит особенность заявленной проблемы.

К психологическим факторам, влияющим на успеваемость студента в ВУЗе можно отнести как индивидуальные способности (интеллект, самоорганизация, умение добиваться поставленных целей, ...), так и показатели, сформированные за период пребывания в университете, - учебная мотивация, свобода творчества, возможность проявить себя, тяга к новым знаниям и т.д. Очевидно, что многие из этих факторов будут коррелировать с успешностью студента. Отсюда следует важный вывод: *на основе данных о психологическом состоянии студента можно сделать прогноз его успеваемости с определенной степенью точности.* Именно это послужило отправной точкой для данной работы. Цель, как упоминалось ранее, состоит в построении модели прогнозирования успеваемости студента по известным психологическим характеристикам.

Проблема, о которой нельзя умолчать, состоит в том, что выявление всех факторов, в действительно влияющих на учебный процесс, представляется очень трудоемкой задачей. Более того, определить точное значение некоторых показателей либо невозможно, либо требует специфичных средств (например, определить уровень интеллекта студента можно только при помощи прохождения специальных тестов; это, в свою очередь, может вызвать негатив со стороны студентов). В связи с этим имеет смысл выделить наиболее информативные факторы и работать исключительно с ними.

В данной работе в качестве факторов, влияющих на успеваемость, были выбраны:

- 1) *интерес к учебе* (коррелирует с мотивацией);
- 2) *сложность обучения в основном учебном заведении* (коррелирует с наличием свободного времени);
- 3) *важность получения высшего образования* (коррелирует с тем, насколько серьезно студент относится к учебе в ВУЗе);

Данные показатели не обязывают студентов к напряженной умственной деятельности, поэтому могут быть легко собраны посредством проведения периодичных мини-опросов.

3. Математическое обоснование

В данной части будет описан ход решения поставленной задачи средствами нечеткой логики.

3.1. Традиционный подход и нечеткая логика

Задача выявления закономерностей в данных и их обобщения – это классическая задача бурно развивающегося направления «Data Science», которое проникает во многие сферы нашей жизни. Действительно, поставленная задача может быть решена традиционными методами машинного обучения или же с помощью применения нейросетевого подхода. Чаще всего для этого требуется:

- 1) иметь репрезентативную выборку исследуемой предметной области
- 2) точечная настройка параметров обучения (выбор алгоритма, подбор различных коэффициентов)

Соблюдение этих условий может привести к хорошим результатам, о чем свидетельствуют многочисленные статьи, посвященные прогнозированию успеваемости студентов с использованием интеллектуального анализа данных (например, [9]).

С другой стороны, гарантировать выполнение этих условий почти никогда невозможно. Например, можно столкнуться с проблемами недостатка данных, «пробелами» и неопределенностями в данных для обучения или же отсутствие технических навыков реализации традиционных подходов (в силу их сложности). Эти и другие обстоятельства приводят к желанию рассмотреть более подходящие альтернативные подходы. Одним из таковых является применение *нечетких математических моделей*.

Нечеткая логика более естественно описывает характер человеческого мышления и ход его рассуждений [3], поскольку основана на нечетких продукционных правилах, чаще всего сформулированных экспертами в определенной предметной области. Это делает данный подход более «прозрачным», гибким и дешевым с точки зрения стоимости различных дополнительных затрат [4].

3.2. Архитектура нечеткой системы

Модель управления, основанная на нечеткой логике, может быть представлена как модель «черного ящика»: получая на вход некоторую информацию (измеренные определенным

образом *входные переменные*), система, используя внутренний механизм нечеткого вывода, выдает ответ, соответствующий *выходным управляющим переменным*.



Рис. 1. Модель «черного ящика»

Внутреннее устройство такой системы («черный ящик») можно интерпретировать как функцию, которая приближенно аппроксимирует реальную зависимость между входной и выходной информацией. Такая функция определяется посредством задания *базы правил системы нечеткого вывода*. Но, чтобы эта функция работала корректно, необходимо ввести нечеткость в поступающую входную информацию (*фаззификация*), а для корректного и интерпретируемого вывода, сделанного системой, - избавиться от нечеткости (*дефаззификация*). При этом требуется, чтобы входные и выходные переменные были представлены как *лингвистические переменные*, каждая из которых обладает определенным множеством *термов*.

Далее рассмотрим введенные понятия нечеткой системы в контексте заявленной задачи.

3.3. Лингвистические переменные

Как упоминалось ранее, в качестве основных факторов для прогнозирования будем использовать такие показатели, как важность образования, интерес студента к учебе и сложность обучения в ВУЗе. По ним модель должна определить успеваемость студента в текущем семестре.

Таким образом, входными лингвистическими переменными являются:

- 1) *важность образования* (importance): оценка того, насколько важно для студента получить высшее образование;

- 2) *интерес к учебе* (interest): оценка желания студента продолжать учиться в текущем ВУЗе;
- 3) *сложность обучения* (difficulty): оценка того, насколько трудно студенту справляться с текущей нагрузкой в университете;

При построении нечеткой модели мы предполагаем, что указанные переменные измеряются в баллах 0 до 10. В качестве термов для переменной «Важность обучения» рассмотрим набор нечетких множеств $T_1 = \{\text{“Бесполезно”, “Нужно исключительно для справки”, “Бывает весьма полезно”, “Важно и необходимо”}\}$. Для переменной «Интерес к учебе» будем использовать множество $T_2 = \{\text{“Отвращение”, “Скучно”, “Привычно”, “Интересно”, “Вдохновляет”}\}$, а для «Сложность обучения» - множество $T_3 = \{\text{“Легко”, “Спокойно”, “Трудновато”, “Сложно”}\}$. Для задания указанных термов воспользуемся *П-образными функциями принадлежности*, поскольку они хорошо учитывают неопределенности типа «около», «примерно», «приблизительно».

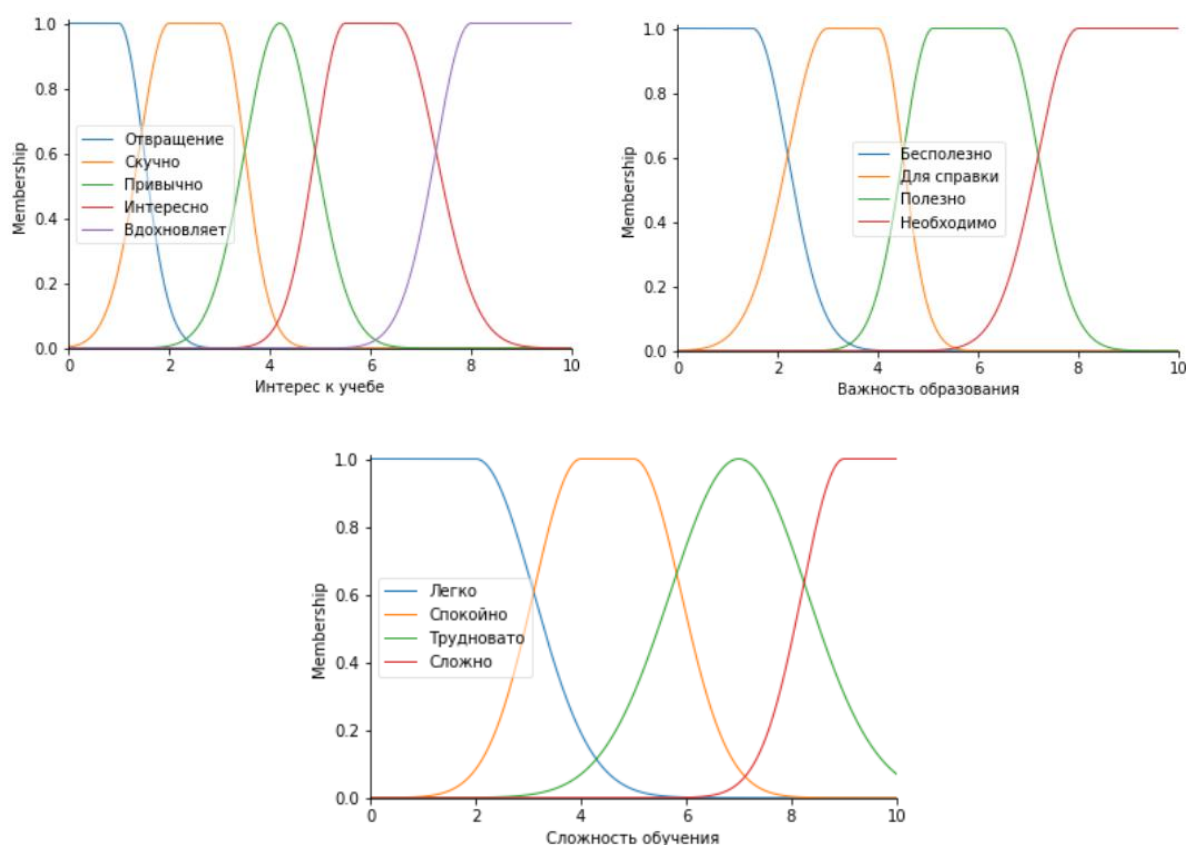


Рис. 2-4. Термы каждой входной лингвистической переменной.

В дальнейшем функции принадлежности и их параметры можно уточнить, но на первоначальном этапе исследования вполне достаточно задать лишь их примерный вид.

В роли выходной лингвистической переменной выступает *успеваемость студента*. Для её описания используем множество термов $T_4 = \{\text{“Плохо”}, \text{“Удовлетворительно”}, \text{“Хорошо”}, \text{“Отлично”}\}$. Диапазон изменения успеваемости: от 2 до 5.

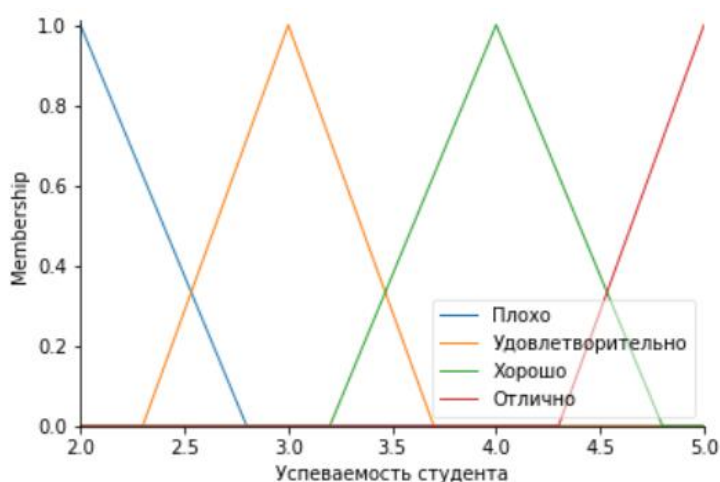


Рис. 5. Термы выходной переменной.

Подразумевается, что выходная переменная характеризует не только средний балл студента, но и в целом описывает оценку результат его учебной деятельности. В связи с этим функции принадлежности выбраны таким образом, чтобы они учитывали неопределенность в «особых» ситуациях (например, когда студент имеет неплохой средний балл, но при этом трудно сказать, что учится он хорошо). В ином случае (если бы успеваемость описывала только средний балл), функции принадлежности для термов можно было бы представить следующим образом:

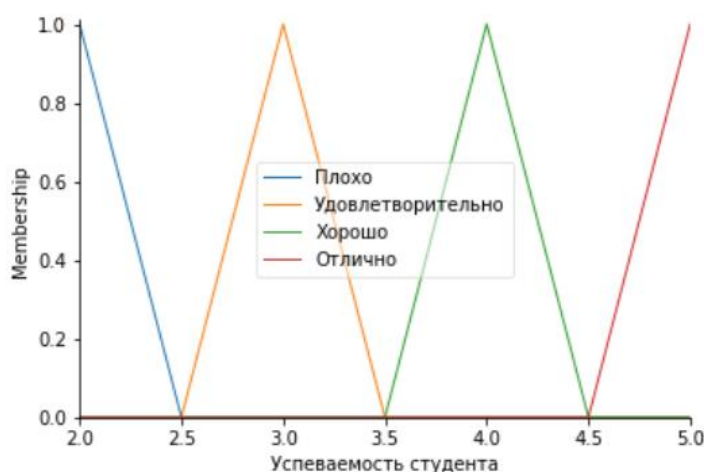


Рис. 6. Термы выходной переменной в случае других функции принадлежности.

3.4. База правил

Существуют различные подходы к формированию правил для нечетких систем. Например, в случае, если в построении системы принимает участие эксперт текущей предметной области, то, используя его знания, можно сформулировать базу правил *априорно*. Это наиболее привлекательный метод, поскольку он не требует сбора данных и хорошо иллюстрирует основную концепцию и цель нечеткой логики.

В случае же, когда эксперта рядом нет (или затраты на привлечение экспертов превосходят наши ожидания), уместно использовать *экспериментальный подход*: собрать некоторую выборку (посредством опроса, например), выделить в ней основные тенденции и, сформулировав их в виде нечетких правил, занести в базу знаний.

В данной работе для формирования базы правил использовался *гибридный метод*. Большая часть правил сформулирована исходя из априорных оценок, но тенденции, наблюдающиеся в контексте рассматриваемой задачи, были определены экспериментальным путем - с помощью проведенного мини-опроса среди студентов группы 19Б11. К таковым закономерностям относятся:

- 1) Повышение показателя важности образования приводит к повышению уровня успеваемости даже в том случае, когда показатель интереса весьма низок.
- 2) Сложность обучения коррелирует с интересом к учебе. Чем выше интерес, тем, как правило, выше и сложность. Однако это не действует в том случае, когда сложность становится невыносимой.
- 3) Как правило, если человеку учиться сложно и интересно, то это значит, что он учится хорошо или отлично. С другой стороны, студенты, учащиеся на уровне удовлетворительно, чаще всего оправдываются тем, что им учиться весьма сложно.

База знаний состоит из 50 правил. В качестве алгоритма логического вывода используется *алгоритм Мамдани*. В роли Т-нормы выступает функция минимума.

3.5. Работа и качество модели

Для анализа построенной модели рассмотрим её работу на некоторых примерах.

- Пусть в качестве входных данных выступают следующие значения: важность оценивается в 3 балла, интерес – в 5 баллов и сложность – в 9 баллов. В таком случае нечеткая модель оценивает успеваемость в 3.389 (рис. 7)
- Пусть важность оценивается в 8 баллов, интерес в 7 баллов, сложность – в 8 баллов. Тогда успеваемость студента, согласно прогнозу модели, равна 4.51 (рис. 8)

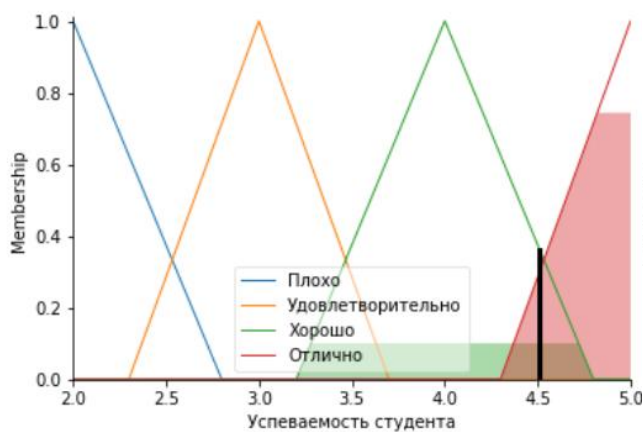


Рис. 7

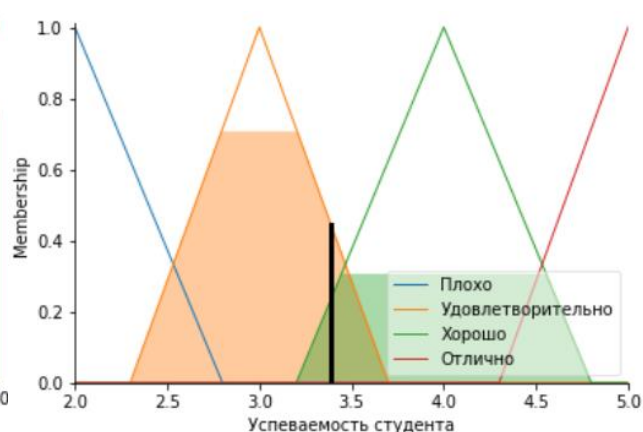


Рис. 8

Для более детального анализа и последующей возможной корректировки параметров системы (функции принадлежности, термов и т.д.) удобно использовать *поверхности нечеткого вывода* (рис. 9-10).

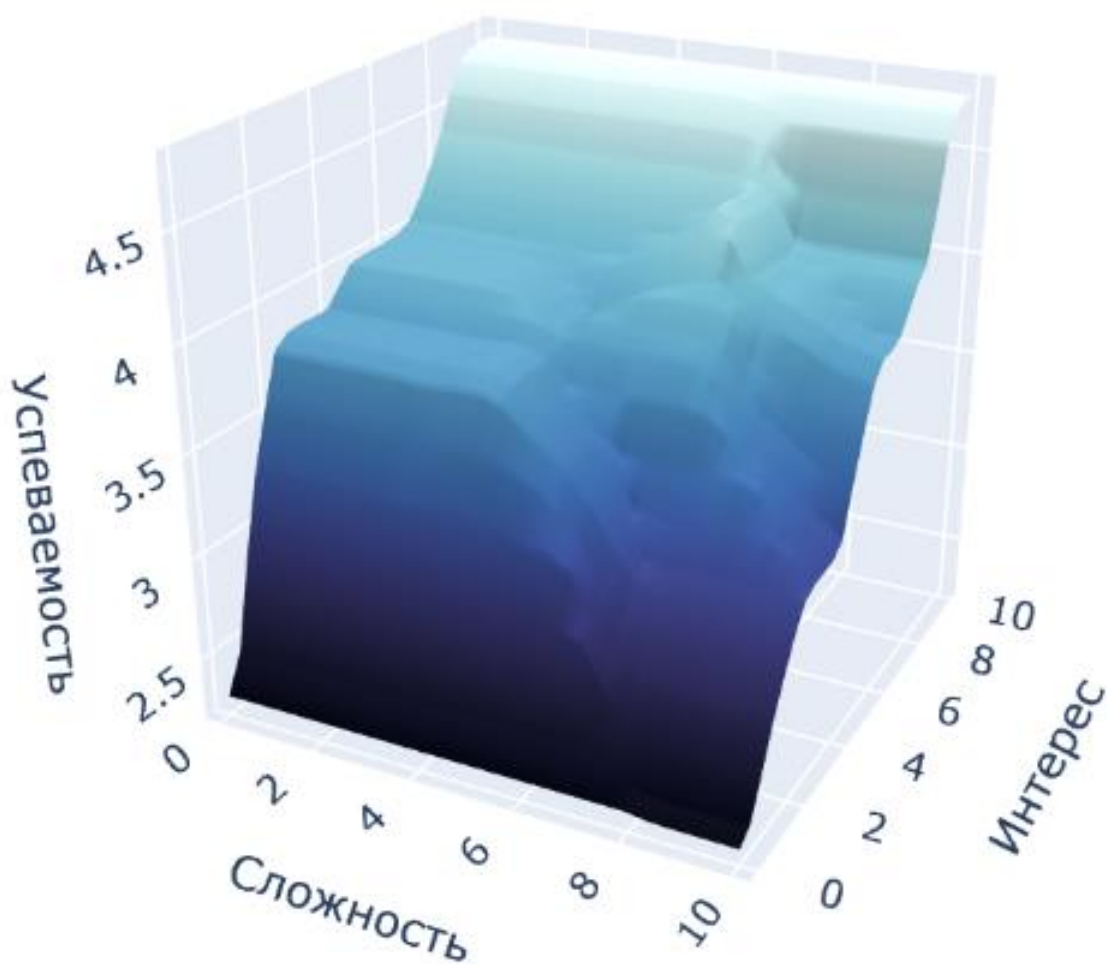


Рис. 9. Поверхность нечеткого вывода для входных переменных «Интерес к учебе» и «Сложность обучения» при фиксированном значении переменной «Важность образования», равном 5.

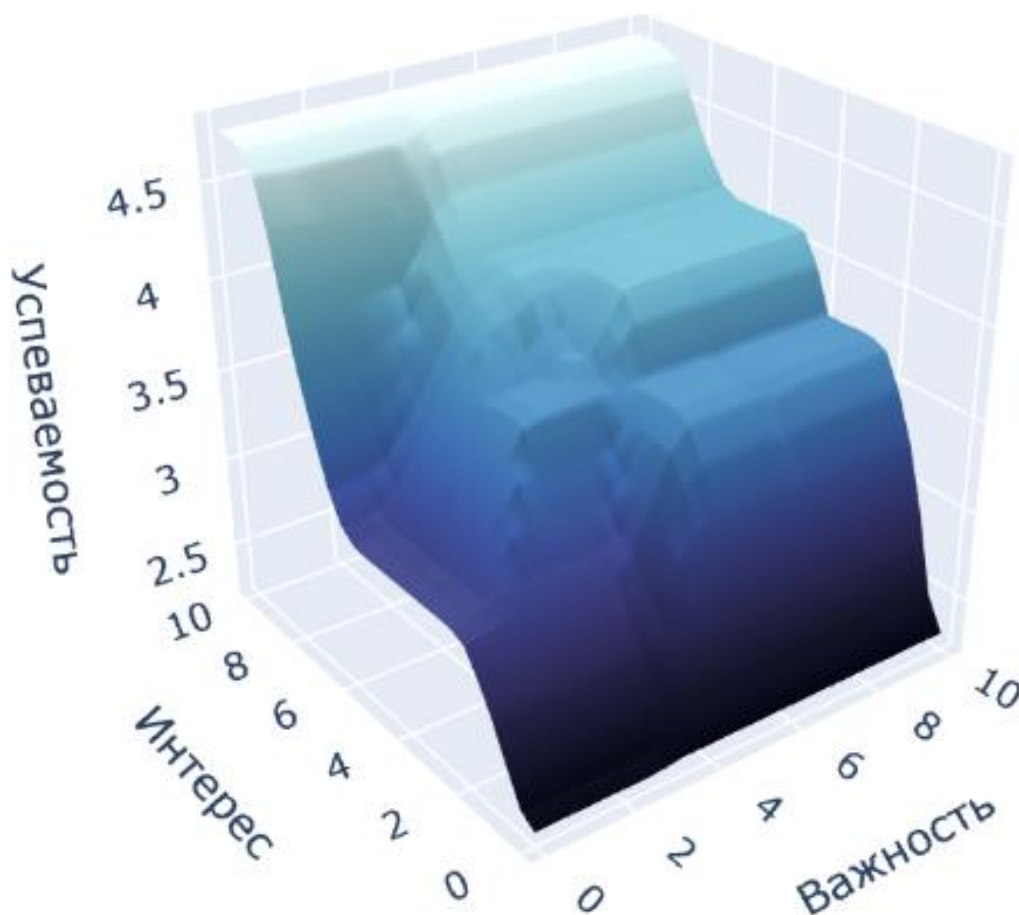


Рис. 10. Поверхность нечеткого вывода для входных переменных «Интерес к учебе» и «Важность образования» при фиксированном значении переменной «Сложность обучения», равном 5.

Поверхности показывают, как меняется управляющая выходная переменная в зависимости от изменения соответствующих входных переменных. В контексте рассматриваемой задачи вполне естественно предполагать, что увеличение таких показателей, как интерес студента к учебе или важность высшего образования для студента, должно приводить к увеличению его успеваемости. В свою очередь увеличение сложности приводит к понижению успеваемости в том случае, если студенту учиться не интересно или для него не так важно получить высшее образование. Представленные выше графики демонстрируют эти и другие тенденции.

3.6. Проблемы и возможные улучшения

Построенная модель может быть значительно улучшена, если принять во внимание особенности построения нечетких систем. Эти особенности во многом представляют собой трудности, которые нужно преодолеть для получения более качественных результатов:

- 1) Субъективный выбор компонентов нечетких систем (термы, лингвистические переменные, правила). Определение компонентов выполняется на основе знаний эксперта. Эти знания могут быть весьма ограничены. Кроме того, использованная для построения системы информация не является полной, поскольку факторов весьма мало по сравнению с действительными показателями. Вариантами решения здесь являются привлечение большего числа экспертов (или же привлечение эксперта более высокого уровня), а также сбор большего числа данных (например, можно добавить такие факторы, как *время, уделяемое учебе*, или *качество отношений с преподавателями* и т.д.). Однако реализация этих идей требует больших затрат.
- 2) Одна из основных проблем нечетких систем – это «правильное» формирование базы знаний. Не существует единой для всех систем метрики качества построенной базы, кроме здравого смысла и анализа результатов в конкретной задаче. Но для адекватной работы модели требуется обеспечить *полноту и непротиворечивость нечетких правил*, что сделать порой весьма затруднительно. Однако это возможно, если подойти к указанному вопросу более тщательно (например, подход, основанный на применении иерархических продукционных правил и регулировании размерности базы знаний, изложен в [8]).
- 3) Ранее уже упоминались традиционные подходы к решению поставленной задачи – использование методов машинного обучения или концепции нейронных сетей. Им в альтернативу были представлены модели, основанные на нечеткой логике. Можно предложить ещё более интересный подход – интеграция нечетких и нейросетевых моделей, основанная на нечеткости и частичной истинности используемых данных [4, глава 3.3]. Эта интеграция может быть реализована в самых различных вариациях, поэтому при исследовании сложных систем как минимум стоит обратить внимание на этот нестандартный подход.

4. Заключение

Формальные модели, решающие задачу прогнозирования успеваемости студента по известным показателям определенных факторов, в большинстве своем основаны на статистическом анализе данных. В данной работе рассмотрен альтернативный подход, идея которого заключается в использовании системы нечеткого вывода. Этот подход более гибкий, поскольку менее требователен к условиям его применения.

В ходе построения нечеткой модели были описаны входные и выходные переменные рассматриваемой задачи и сформулирована база знаний, состоящая из эвристических правил. Для реализации системы использовалась библиотека skfuzzy языка Python [6]. Для визуализации поверхностей использовалась библиотека plotly [7]. Соответствующий код может быть найден по адресу: <https://github.com/AnatolyAdamovich/fuzzy-logic-model>

Результаты проведенного исследования приводят к мысли о том, что при более качественной настройке параметров системы, более тщательном выборе функций принадлежности и прочих возможных улучшениях удастся достичь ещё более высокой точности модели и повысить качество прогнозирования.

5. Список литературы

1. **Зяблецев П.А.** Прогнозная модель для оценки успеваемости студентов университета по итогам текущего обучения (Магистерская диссертация, ТПУ. 2020г.)
2. **Смирнов С. Д.** Психологические факторы успешной учебы студентов вуза.
3. **А. В. Леоненков.** Нечеткое моделирование в среде MATLAB и fuzzyTECH.
4. **В.В. Борисов, В. В. Круглов, А. С. Федулов.** Нечеткие модели и сети [2-е издание].
5. **А. Г. Броневи́ч, А. Е. Лепский.** Нечеткие модели анализа данных и принятия решений [Издательский дом Высшей школы экономики, 2022].
6. **skfuzzy.** <https://scikit-fuzzy.readthedocs.io/en/latest/>
7. **plotly.** <https://plotly.com/python/>
8. **М. А. Сергиенко.** Методы проектирования нечеткой базы знаний. (Воронежский государственный университет).
9. **Muhammad Imran, Shahzad Latif Latif, Danish Mahmood, Musammad Shah.** Student Academic Perfomance Prediction using Supervised Learning Techniques. [2019, *Internation Journal of Emerging Technololgies in Learning*]