

# Предсказание вязкости жидкостей с помощью методов машинного обучения на основе параметров уравнения состояния

Панов Михаил Федорович  
Руководитель: Писарев Василий Вячеславович

Москва --- 2025

## **Задание на ВКР**

(Текст задания на ВКР)

## **Аннотация**

(Аннотация на русском языке)

## **Abstract**

(Аннотация на английском языке)

# Содержание

<b>Введение</b>	<b>5</b>
<b>1 Обзор литературы и постановка задачи</b>	<b>5</b>
1.1 Свойства вязкости и ее роль в инженерных задачах . . . . .	5
1.1.1 Вязкость в промышленных приложениях . . . . .	5
1.1.2 Значение точного прогнозирования вязкости . . . . .	5
1.2 Обзор подходов к предсказанию вязкости . . . . .	5
1.2.1 Ограничения эмпирических формул . . . . .	5
1.2.2 Модель MYS . . . . .	6
1.2.3 Уравнение состояния CP-PC-SAFT . . . . .	7
1.2.4 Параметры веществ в CP-PC-SAFT . . . . .	8
1.2.5 Трудности машинного обучения . . . . .	9
1.2.6 Символьная регрессия . . . . .	10
1.3 Цель и задачи исследования . . . . .	11
1.3.1 Цель исследования . . . . .	11
1.3.2 Задачи исследования . . . . .	11
1.4 Выводы по главе . . . . .	12
<b>2 Сбор и обработка данных</b>	<b>13</b>
2.1 Работа с базой данных . . . . .	13
2.1.1 Источник данных: база ThermoML . . . . .	13
2.1.2 ThermoPyL для чтения данных . . . . .	13
2.2 Выбор веществ и критерии включения . . . . .	13
2.3 Выбор веществ и критерии включения . . . . .	14
2.4 Добавление параметров уравнения CP-PC-SAFT . . . . .	14
2.5 Вычисление производных величин . . . . .	14
2.5.1 Мольный объем . . . . .	14
2.5.2 Избыточная энтропия . . . . .	14
2.6 Выводы по главе . . . . .	14
<b>3 Построение моделей и генерация признаков</b>	<b>15</b>
3.1 Автоматическая генерация признаков . . . . .	15
3.2 Отбор признаков . . . . .	15
3.3 Символьная регрессия . . . . .	15
3.4 Проверка значимости избыточной энтропии . . . . .	15
3.5 Выводы по главе . . . . .	15
<b>4 Обучение моделей и анализ результатов</b>	<b>16</b>
4.1 Обучение моделей ML . . . . .	16
4.2 Сравнение подходов . . . . .	16
4.3 Сравнение с моделью MYS . . . . .	16
4.4 Выводы по главе . . . . .	16
<b>Заключение</b>	<b>17</b>
<b>Список литературы</b>	<b>18</b>

# Введение

(Введение: актуальность, цель, задачи, объект, структура)

## 1. Обзор литературы и постановка задачи

### 1.1. Свойства вязкости и ее роль в инженерных задачах

#### 1.1.1. Вязкость в промышленных приложениях

Вязкость — ключевой реологический параметр, определяющий сопротивление жидкости деформации при сдвиге. Ее значение напрямую влияет на гидродинамические характеристики систем, включая распределение давления, скорости потока и теплопередачу. В ряде инженерных процессов, таких как транспортировка нефти и газа, проектирование теплообменников, химическая переработка и производство смазочных материалов, точное знание вязкости определяет эффективность и безопасность эксплуатации оборудования.

В многофазных или сложных многокомпонентных системах, например, в синтетических нефтяных смесях, вязкость влияет на фазовое поведение, устойчивость эмульсий и общую производительность процессов. В микрофлюидике и фармацевтике контроль вязкости необходим для предсказуемого управления потоками и дозированием веществ.

#### 1.1.2. Значение точного прогнозирования вязкости

Точное моделирование вязкости представляет собой одну из ключевых задач в термодинамическом моделировании жидкостей. На практике невозможно измерить вязкость для всех возможных условий и составов, особенно в системах, содержащих десятки или сотни компонентов, как, например, в нефтяной промышленности. В таких случаях используются предсказательные модели, основанные на молекулярных или термодинамических представлениях, которые позволяют получить значения вязкости при отсутствии экспериментальных данных.

Среди современных подходов особое внимание уделяется моделям, связывающим вязкость с остаточной энтропией и другими термодинамическими величинами, так как они обеспечивают высокую переносимость и физическую обоснованность. Связь между вязкостью и избыточной энтропией подмечали во многих работах, например [1]. Особенно актуальны такие модели для расчетов в широком диапазоне температур и давлений, а также при переходе к сверхкритическим и конденсированным фазам.

Таким образом, разработка универсальных моделей вязкости, не требующих индивидуальной подгонки параметров, является приоритетной задачей в области физико-химического моделирования сложных жидкостей.

## 1.2. Обзор подходов к предсказанию вязкости

### 1.2.1. Ограничения эмпирических формул

Большинство классических подходов к описанию вязкости основаны на эмпирических уравнениях, содержащих подобранные коэффициенты для конкретных веществ или их классов. Такие модели, как правило, работают в узком диапазоне температур и давлений и требуют предварительных измерений вязкости. Это делает их трудноприменимыми для новых

или редких веществ, особенно в условиях, когда экспериментальные данные недоступны или ограничены. (for example)

### 1.2.2. Модель MYS

Одной из наиболее универсальных и известных современных моделей для описания динамической вязкости жидкостей в широком диапазоне условий является модель, предложенная Ильей Полищуком в 2015 году — так называемая *Modified Yarranton–Satyro (MYS)* модель [2]. Ее основное назначение — обеспечить точное предсказание вязкости для широкого спектра органических жидкостей, включая углеводороды, в диапазонах температур от тройной точки до высоких значений, а также при изменяющемся давлении.

**Предпосылки модели.** Изначально модель Yarranton–Satyro (YS) создавалась для описания вязкости как функции плотности, температуры и параметров молекулярного строения. Polishuk модифицировал ее, чтобы избежать необходимости экспериментальной подгонки и позволить использовать только параметры, выводимые из уравнений состояния (в частности, SAFT). Основной целью стало создание аналитической корреляции, не требующей знания вязкости как входного параметра, но использующей только предсказуемые или известные характеристики вещества.

**Структура модели.** Модель MYS представляет собой сложную аналитическую формулу, включающую молекулярные параметры, полученные из теории SAFT:

$$\eta = 0.1 \left( \exp \left( \frac{c_1 \sqrt{m} + \ln \left( 1 + \frac{M_w^4}{c_2 v^4 m^3} \right) \ln \left( \frac{T}{T_{\text{tr}} + 120} \right)}{\exp \left( \frac{1.04v}{N_{\text{av}} m \sigma^3} \right) \exp \left( \frac{c_3 P}{\epsilon_k \sqrt{-\frac{dP}{dv} \cdot \sqrt{mv}}} \right) - 1} \right) - 1 \right) + \eta_0$$

где:

- $\eta$  — динамическая вязкость (Па·с)
- $\eta_0$  — вязкость идеального газа (обычно берется как постоянная или ноль)
- $m$  — молярная масса
- $M_w$  — молекулярная масса
- $v$  — мольный объем
- $T$  — температура
- $T_{\text{tr}}$  — температура тройной точки
- $P$  — давление
- $\epsilon_k$  — параметр межмолекулярного взаимодействия
- $dP/dv$  — производная давления по объему (рассчитанная из уравнения состояния)
- $\sigma$  — эффективный диаметр молекулы
- $N_{\text{av}}$  — число Авогадро
- $c_1 = 0.27$ ,  $c_2 = 2.5 \cdot 10^{11}$ ,  $c_3 = 2.1$  — эмпирические коэффициенты (подобраны автором)

**Особенности модели.** - Формула MYS не требует априорного знания вязкости, но использует величины, производные из термодинамических функций, так как она связана с параметрами уравнения состояния CP-PC-SAFT. - Автор демонстрирует среднюю ошибку модели на уровне менее 10% для широкой выборки жидкостей. - Подход особенно эффективен для органических и углеводородных жидкостей.

**Применение в данной работе.** Модель MYS была выбрана в качестве основного ориентира для сравнения, поскольку:

- Она является одной из наиболее точных универсальных формул, не требующих подбора индивидуальных коэффициентов для каждого вещества.
- Она опирается на термодинамически интерпретируемые параметры, аналогично нашему подходу.
- Ее точность считается приемлемой для инженерных и прикладных задач.
- Она сохраняет свою точность на достаточно большом диапазоне давлений и температур

В настоящем исследовании модели машинного обучения и символьной регрессии сравниваются по точности предсказания вязкости с моделью MYS. Наши модели демонстрируют в ряде случаев ошибку до 2% — значительно ниже, чем у MYS — что позволяет рассматривать их как более точную альтернативу в условиях, когда имеются данные для обучения модели.

### 1.2.3. Уравнение состояния CP-PC-SAFT

Основное уравнение состояния PC-SAFT выражается через вклад свободной энергии Гельмгольца  $a$ :

$$A(v, T) = A^{\text{id}} + A^{\text{hs}} + A^{\text{chain}} + A^{\text{disp}} + A^{\text{assoc}} \quad (1)$$

Где:

- $A^{\text{id}}$  — вклад идеального газа;
- $A^{\text{hs}}$  — вклад жёстких сфер;

$$A^{\text{HS}} = RT \frac{m}{\zeta_0} \left( \frac{3\zeta_1\zeta_2}{1-\zeta_3} + \frac{\zeta_2^2}{\zeta_3(1-\zeta_3)^2} \right) + \left( \frac{\zeta_2^2}{\zeta_3^2} - \zeta_0 \right) \ln[1-\zeta_3] \sqrt{\frac{d^3(\zeta_3-1)}{\zeta_3\sigma^3-d^3}}$$

- $A^{\text{chain}}$  — вклад цепных взаимодействий;

$$A^{\text{chain}} = RT \sum_i x_i x_j (1 - m_{ij}) \ln[g_{ij}(d_{ij})]^{\text{hs}}$$

- $A^{\text{disp}}$  — вклад дисперсионных сил;

$$A^{\text{disp}} = -RN_{\text{Av}} \left( \frac{2\pi(\epsilon/k)m^2\sigma^3}{\nu} I_1 + \frac{\pi(\epsilon/k)^2 m^3 \sigma^3}{\nu T \left( 1 + \frac{m(8\zeta_3-2\zeta_3^2)}{(1-\zeta_3)^4} + \frac{(1-m)(20\zeta_3-27\zeta_3^2+12\zeta_3^3-2\zeta_3^4)}{((1-\zeta_3)(2-\zeta_3))^2} \right)} I_2 \right) \quad (2)$$

- $A^{\text{assoc}}$  — вклад ассоциативных взаимодействий.
- $d$  — эффективный диаметр молекулы, связанный с  $\sigma$  через  $\theta$ .

$$d = \theta * \sigma$$

- $\sigma$  — характерный размер молекулы (диаметр сферического сегмента).
- $\theta$  — коэффициент масштабирования длины связи.

$$\theta = \frac{1 + 0.2977(k/\epsilon)T}{1 + 0.33163(k/\epsilon)T + 0.0010477(k/\epsilon)^2T^2}$$

- $\zeta_k$  — параметр упаковки молекул.

$$\zeta_k = \frac{\pi N_{av}}{6\nu} \sum_i x_i m_i \sigma_i^{dk}$$

- $\epsilon/k_B$  — приведённая энергия взаимодействия между частицами.
- $N_{Av}$  — число Авогадро.
- $m$  — количество сегментов в молекуле.
- $x_i$  — мольная доля компонента  $i$ .
- $P$  — давление.
- $T$  — температура.
- $V_m$  — мольный объём.
- $d_{ij}$  — средний эффективный диаметр молекул  $i$  и  $j$ .
- $g_{ij}$  — функция радиальной корреляции молекул  $i$  и  $j$ .
- $\nu$  — молярный объём системы.
- $I_1, I_2$  — интегралы в дисперсионном вкладе.

#### 1.2.4. Параметры веществ в CP-PC-SAFT

В рамках CP-PC-SAFT каждое вещество описывается тремя основными параметрами:

- $m$  — число сегментов в молекуле;
- $\sigma$  — эффективный диаметр сегмента молекулы;
- $\epsilon$  — энергия взаимодействия между сегментами.

Эти параметры определяются через критическую точку и температуру кипения вещества, а не через эмпирическую подгонку, что делает модель более универсальной.

В модели CP-PC-SAFT параметры вещества корректируются так, чтобы соответствовать следующим условиям [3]:

$$\left(\frac{\partial P}{\partial v}\right)_{T_c} = 0, \quad \left(\frac{\partial^2 P}{\partial v^2}\right)_{T_c} = 0 \quad (3)$$

$$P_c = P_{c, \text{exp}} \quad (4)$$

$$\rho_{\text{liq, triple}} = \rho_{\text{liq, triple, exp}} \quad (5)$$

Здесь:

- $P_c$  — давление в критической точке;
- $T_c$  — критическая температура;
- $\rho_{\text{liq, triple}}$  — жидкостная плотность в тройной точке.

Условиями для поиска параметров являются критическая температура, критическое давление и плотность жидкости при температуре плавления, соответствующие экспериментальным значениям. Необходимость знания значений всего трех экспериментальных величин для вычисления всех параметров вещества делает модель CP-PC-SAFT крайне привлекательной в вопросах предсказания свойств новых веществ и смесей на большом диапазоне условий.

### 1.2.5. Трудности машинного обучения

Применение методов машинного обучения (ML) к задачам физико-химического моделирования, включая предсказание вязкости жидкостей, в последние годы стало популярным направлением (for example). Такие методы, как правило, позволяют автоматически выявлять сложные зависимости в данных и достигать высокой точности без необходимости ручного выбора аналитических формул. Тем не менее, несмотря на прогресс в области ML, существует ряд фундаментальных причин, по которым машинное обучение нельзя считать универсальным решением (панацеей) для предсказания реологических свойств веществ.

**1. Ограниченная интерпретируемость моделей.** Большинство моделей машинного обучения, включая градиентный бустинг, нейронные сети и случайные леса, представляют собой черные ящики, в которых интерпретация внутренней логики и значимости отдельных признаков затруднена. Это особенно критично в контексте задач молекулярной термодинамики, где важна не только точность, но и возможность объяснить результат через физически обоснованные параметры. В противоположность этому, аналитические формулы, такие как MYS, или, потенциально, символьная регрессия (используемая в данной работе) позволяют сохранять интерпретируемость и анализировать вклад каждой переменной.

**2. Риск переобучения и проблемы обобщения.** Как показано в ходе данной работы, даже простые модели, такие как линейная регрессия или случайный лес, демонстрируют высокий риск переобучения при специальном разбиении выборки. В частности, точность существенно снижается при валидации на новых веществах, не представленных в обучающей выборке. Особенно ярко это проявляется в случае веществ, отличающихся по агрегатному состоянию — например, бутана. Это указывает на ограниченную способность моделей ML к интерполяции за пределами обучающих данных.



**3. Зависимость от качества и полноты данных.** Модели машинного обучения требуют большого количества однородных и хорошо размеченных данных. В случае термодинамических величин, таких как вязкость, часто доступны только ограниченные экспериментальные измерения, сделанные в различающихся условиях и с неоднородной точностью. Это может как ограничивать возможности ML, делая необходимым использование большего количества уникальных производных признаков, полученных, например, из уравнений состояния, так и приводить к переобучению из-за попыток модели подстроиться к условиям конкретного эксперимента.

**4. Отсутствие устойчивости к физическим границам.** Некоторые модели ML могут выдавать предсказания, нарушающие физические ограничения — например, отрицательную вязкость или аномальные значения при экстремальных температурах и давлениях. Это особенно заметно в моделях, не обученных на таких данных. В отличие от этого, аналитические формулы и интерпретируемые модели могут быть сконструированы с учетом физических асимптотик и ограничений.

**5. Практические требования к применению.** Для внедрения моделей в инженерную практику важны не только высокая точность, но и компактность, воспроизводимость и простота вычислений. Модели машинного обучения, особенно глубокие нейросети, могут быть ресурсоемкими и плохо воспроизводимыми. Символьные формулы или регрессии, напротив, могут быть легко реализованы даже в простых инженерных расчетах.

Таким образом, несмотря на мощность и гибкость современных методов машинного обучения, они должны использоваться с осторожностью, особенно в задачах с физико-химическим контекстом. Наиболее перспективным направлением видится интеграция ML с физически обоснованными подходами — в частности

#### 1.2.6. Символьная регрессия

Символьная регрессия (symbolic regression, SR) — это подход к построению моделей, при котором алгоритм не просто находит численные параметры в заданной структуре формулы как в линейной регрессии, а ищет саму структуру уравнения, наиболее точно описывающего зависимость между входными и выходными переменными. В отличие от традиционных методов машинного обучения, символьная регрессия формирует аналитическое выражение в виде комбинаций математических операций и входных признаков.

**Мотивация применения.** Основным преимуществом SR является высокая интерпретируемость получаемых моделей. Вместо подбора большого количества слабоинтерпретируемых коэффициентов, как в случае с нейросетями, модель находит короткие выражения, состоящие из качественно различающихся блоков (операторов). Каждое уравнение, сформированное таким способом, может быть проанализировано, сопоставлено с физическими законами, а также непосредственно использовано в инженерных расчетах. Это особенно важно при работе с термодинамическими и молекулярными свойствами, где значение имеет не только точность, но и физический смысл модели.

В рамках настоящего исследования SR использовалась для поиска наиболее компактных и точных выражений, связывающих вязкость с параметрами уравнения состояния CP-PC-SAFT и производными термодинамическими величинами, включая мольный объем, избыточную энтропию и производную давления по объему.

**Используемый инструмент: PySR.** Для реализации символьной регрессии применялась современная библиотека PySR (Python Symbolic Regression) [4]. Эта библиотека сочетает возможности языка Python для настройки задач и быстроедействие языка Julia, где реализована основная эволюционная оптимизация. PySR использует алгоритмы на основе генетиче-

ского программирования, направленные на поиск выражений с оптимальным компромиссом между точностью и сложностью.

Основные особенности PySR:

- Многофункциональная система операторов: поддерживаются стандартные математические функции (логарифм, экспонента, степень и др.).
- Поддержка многоцелевой оптимизации: минимизация ошибки при контроле за длиной формулы.
- Устойчивость к переобучению: регуляризация сложности встроена в процесс поиска.
- Высокая производительность благодаря параллельному выполнению на CPU и GPU.

**Результаты применения.** В ходе экспериментов символьная регрессия позволила получить компактные формулы с ошибкой, сравнимой с ошибкой модели MYS. Таким образом, символьная регрессия показала себя как перспективный инструмент, позволяющий извлекать простые закономерности. Ее применение может дополнить использование других моделей и усилить акцент на физическую обоснованность результатов.

### 1.3. Цель и задачи исследования

#### 1.3.1. Цель исследования

Разработка метода прогнозирования вязкости жидкостей на основе параметров уравнения состояния CP-PC-SAFT с точностью, превосходящей существующую модель MYS.

#### 1.3.2. Задачи исследования

Для реализации поставленной цели были намечены следующие задачи:

##### а) Сбор и подготовка экспериментальных данных

- Получение доступа к данным термодинамических параметров веществ.
- Формирование базы данных свойств жидкостей и газов на основе ThermoML.
- Фильтрация выборки для получения только интересующих веществ с замеренными температурой, давлением и вязкостью.
- Предварительная обработка данных для приведения интересующих параметров к единому формату.

##### б) Расчет производных параметров

- Вычисление мольного объема по известным температуре и давлению.
- Определение избыточной энтропии  $S^{\text{ex}}$  по известным мольному объему и температуре через производную избыточной энергии Гельмгольца по температуре.
- Вычисление вязкости идеального газа.

##### в) Генерация и отбор новых признаков, проверка их значимости

- Создание производных признаков, отражающих физические закономерности вязкости.

- Оценка их статистической значимости для модели, и отбор наиболее полезных.
- Валидация роли различных признаков в прогнозировании вязкости, в том числе избыточной энтропии.

#### г) Обучение, сравнение и оптимизация моделей машинного обучения

- Построение и проверка эффективности различных моделей ML, таких как линейная регрессия, градиентный бустинг, случайный лес и нейросетевые методы.
- Оптимизация гиперпараметров для повышения точности наиболее перспективных моделей.
- Использование символьной регрессии (PySR) для поиска интерпретируемых аналитических зависимостей.

#### д) Сравнение разработанной модели с существующей моделью MYS

- Оценка точности предложенного метода на тестовых данных.
- Анализ преимуществ и недостатков разработанных моделей по сравнению с MYS.
- Формирование рекомендаций по дальнейшему развитию метода прогнозирования вязкости жидкостей.

### 1.4. Выводы по главе

В первой главе был проведён обзор современных представлений о вязкости как важном термодинамическом параметре и существующих подходах к её моделированию. Показано, что вязкость оказывает существенное влияние на поведение жидких и многофазных систем в различных инженерных приложениях — от нефтепереработки до микрофлюидики.

Рассмотрены эмпирические формулы, применяемые для расчёта вязкости, а также универсальные аналитические модели, такие как MYS (Modified Yarranton–Satyro). Несмотря на широкую применимость первых, они имеют ограничения в точности при отсутствии экспериментальных данных для подгонки или при переходе к новым веществам. В свою очередь, MYS не всегда даёт достаточные для прикладных задач оценки ввиду своей универсальности.

Особое внимание было уделено уравнению состояния CP-PC-SAFT, которое позволяет получить необходимые параметры вещества без тщательной подгонки, опираясь лишь на критические и тройные точки. Это делает его особенно подходящим для предсказательных моделей, работающих на широком спектре условий и веществ, таких как MYS или разрабатываемая в рамках данной работы.

Кроме того, рассмотрены возможности машинного обучения и его ограничения в задачах физико-химического моделирования, включая проблемы интерпретируемости, переобучения и нарушения физических ограничений. На этом фоне была обоснована целесообразность использования символьной регрессии и системы генерации новых признаков как методов, сочетающих гибкость ML с интерпретируемостью и физической обоснованностью аналитических формул.

Таким образом, глава обосновывает научную и практическую значимость задачи предсказания вязкости на основе параметров уравнения состояния и подтверждает пользу разработки нового подхода, сочетающего машинное обучение и следование физическим принципам. Далее в работе рассматривается реализация этого подхода на практике.

## 2. Сбор и обработка данных

### 2.1. Работа с базой данных

#### 2.1.1. Источник данных: база ThermoML

Для получения экспериментальных данных о вязкости в данной работе использовалась база данных **ThermoML**, разработанная Национальным институтом стандартов и технологий США (NIST). ThermoML представляет собой машиночитаемый формат на основе XML, предназначенный для хранения и распространения термодинамических и физико-химических данных, включая фазовые равновесия, плотности, давления, температуры и вязкости чистых веществ и смесей.

Одной из ключевых особенностей ThermoML является строго стандартизированная структура описания данных, что позволяет автоматизировать их извлечение и минимизировать необходимость ручной предобработки. Это особенно важно при формировании выборок для задач машинного обучения, где требуется согласованность единиц измерения и форматов.

#### 2.1.2. ThermoPyL для чтения данных

Для взаимодействия с базой ThermoML была использована специализированная Python-библиотека **ThermoPyL**. Данная библиотека предоставляет инструменты для:

- автоматического чтения и парсинга XML-файлов ThermoML;
- фильтрации данных по фазе и составу;
- преобразования данных в формат `pandas.DataFrame` для последующей обработки в Python;
- агрегации данных по условиям экспериментов;
- экспорта результатов в формат CSV.

Использование библиотеки ThermoPyL позволило автоматизировать процесс подготовки данных, не углубляясь в анализ и обработку XML файлов из ThermoML.

## 2.2. Выбор веществ и критерии включения

Для анализа были выбраны чистые вещества, относящиеся к классу насыщенных углеводородов. Основным критерием отбора было наличие значительного объема экспериментальных данных по вязкости в базе ThermoML. В финальную выборку вошли восемь веществ:

- Бутан: получено 311 экспериментальных точек из 2 работ [5, 6]
- Пентан: получено 50 экспериментальных точек из 3 работ [7—9]
- Гексан: получено 140 экспериментальных точек из 25 работ [6, 8, 10—32]
- Гептан: получено 443 экспериментальных точек из 28 работ [8, 10, 14, 15, 19, 21, 25, 30, 33—52]
- Октан: получено 380 экспериментальных точек из 32 работ [6—8, 14, 15, 17—19, 21, 26, 30, 33, 35, 37, 38, 40, 43, 44, 46, 52—64]

- Нонан: получено 115 экспериментальных точек из 13 работ [7, 8, 14, 15, 33, 35, 38, 44, 47, 60, 65—67]
- Декан: получено 242 экспериментальных точек из 25 работ [7, 8, 14, 15, 17, 18, 26, 31, 33, 35, 38, 44, 54, 57, 60, 62—64, 66, 68—73]
- Додекан: получено 425 экспериментальных точек из 25 работ [7, 14, 15, 27, 39, 44, 53, 54, 57, 65, 69, 74—87]

Для включения записи в выборку требовалось наличие всех трёх параметров:

- температуры ( $T$ , К),
- давления ( $P$ , кПа),
- вязкости ( $\eta$ , Па·с).

Записи, в которых отсутствовала хотя бы одна из этих величин, исключались. Дополнительная фильтрация по фазе не производилась, поскольку, например, для бутана полезны данные как в жидкой, так и в газообразной фазе. Каждая выборка сохранялась в виде отдельного CSV-файла для дальнейшей обработки. В результате было собрано более 2000 уникальных записей.

## 2.3. Выбор веществ и критерии включения

## 2.4. Добавление параметров уравнения CP-PC-SAFT

## 2.5. Вычисление производных величин

### 2.5.1. Мольный объем

### 2.5.2. Избыточная энтропия

Основная идея заключается в том, что логарифм безразмерной вязкости жидкости может быть представлен как функция избыточной энтропии вещества, нормализованной по постоянной Больцмана. Таким образом, в основу модели заложена физически обоснованная связь между вязкостью и степенью упорядоченности молекулярной структуры жидкости, выраженной через избыточную энтропию:

$$\ln \left( \frac{\eta}{\eta_0} \right) = A + B \cdot \frac{S^{\text{ex}}}{k_B} + C \cdot \left( \frac{S^{\text{ex}}}{k_B} \right)^2 + \dots$$

где: -

$\eta$  — динамическая вязкость жидкости, -

$\eta_0$  — характеристическая (референсная) вязкость, -  $S^{\text{ex}}$  — избыточная энтропия, -  $A, B, C$  — подбираемые параметры.

## 2.6. Выводы по главе

### **3. Построение моделей и генерация признаков**

#### **3.1. Автоматическая генерация признаков**

#### **3.2. Отбор признаков**

#### **3.3. Символьная регрессия**

#### **3.4. Проверка значимости избыточной энтропии**

#### **3.5. Выводы по главе**

## **4. Обучение моделей и анализ результатов**

### **4.1. Обучение моделей ML**

### **4.2. Сравнение подходов**

### **4.3. Сравнение с моделью MYS**

### **4.4. Выводы по главе**

## **Заключение**

(Основные выводы и перспективы)



## Список литературы

- [1] Malyanah Binti Mohd Taib и J. P. Martin Trusler. «Residual entropy model for predicting the viscosities of dense fluid mixtures». В: *The Journal of Chemical Physics* 152.16 (2020), с. 164104. DOI: 10.1063/5.0002242. URL: <https://pubs.aip.org/aip/jcp/article/152/16/164104/198785/Residual-entropy-model-for-predicting-the>.
- [2] Ilya Polishuk. «A Modeling Framework for Predicting and Correlating Viscosities of Liquids in Wide Range of Conditions». В: *Industrial & Engineering Chemistry Research* 54.27 (2015), с. 6999—7003. DOI: 10.1021/acs.iecr.5b01468. URL: <https://doi.org/10.1021/acs.iecr.5b01468>.
- [3] Ilya Polishuk. «Standardized critical point-based numerical solution of statistical association fluid theory parameters: the perturbed chain-statistical association fluid theory equation of state revisited». В: *Industrial & Engineering Chemistry Research* 53.36 (2014), с. 14127—14141.
- [4] Miles Cranmer. *PySR: Fast & Interpretable Symbolic Regression in Python and Julia*. 2023. arXiv: 2305.01582 [cs.LG].
- [5] Sebastian Herrmann и Eckhard Vogel. В: *J. Chem. Eng. Data* 60 (2015), с. 3703—3720. DOI: 10.1021/acs.jced.5b00654.
- [6] A. Kumagai, D. Tomida и C. Yokoyama. «Measurements of the Liquid Viscosities of Mixtures of n-Butane, n-Hexane, and n-Octane with Squalane to 30 MPa». В: *International Journal of Thermophysics* 27 (2006), с. 376—393. DOI: 10.1007/s10765-006-0053-2.
- [7] Alfredo Pimentel-Rodas, Luis A. Galicia-Luna и José J. Castro-Arellano. В: *J. Chem. Eng. Data* 62 (2017), с. 3946—3957. DOI: 10.1021/acs.jced.7b00650.
- [8] Hossein Iloukhani, Mahdi Rezaei-Sameti и Jalal Basiri-Parsa. В: *The Journal of Chemical Thermodynamics* 38 (2006), с. 975—982. DOI: <https://doi.org/10.1016/j.jct.2005.10.011>.
- [9] H. Iloukhani и M. Rezaei-Sameti. В: *J. Chem. Eng. Data* 50 (2005), с. 1928—1931. DOI: 10.1021/je0501944.
- [10] Alfredo Pimentel-Rodas, Luis A. Galicia-Luna и José J. Castro-Arellano. В: *J. Chem. Eng. Data* 61 (2016), с. 45—55. DOI: 10.1021/acs.jced.5b00152.
- [11] Waqar Ahmad и др. В: *J. Chem. Eng. Data* 64 (2019), с. 459—470. DOI: 10.1021/acs.jced.8b00589.
- [12] Pedro Morgado и др. В: *Fluid Phase Equilibria* 358 (2013), с. 161—165. DOI: <https://doi.org/10.1016/j.fluid.2013.07.060>.
- [13] S.G. Mohammed Hussain и др. В: *Fluid Phase Equilibria* 476 (2018), с. 139—146. DOI: <https://doi.org/10.1016/j.fluid.2018.08.001>.
- [14] Adel S. Al-Jimaz и др. В: *The Journal of Chemical Thermodynamics* 37 (2005), с. 631—642. DOI: <https://doi.org/10.1016/j.jct.2004.09.021>.
- [15] J.G. Baragi и др. В: *The Journal of Chemical Thermodynamics* 38 (2006), с. 75—83. DOI: <https://doi.org/10.1016/j.jct.2005.03.024>.
- [16] A. Rodríguez и др. В: *The Journal of Chemical Thermodynamics* 38 (2006), с. 505—519. DOI: <https://doi.org/10.1016/j.jct.2005.07.008>.

- [17] Gyan Prakash Dubey, Monika Sharma и Neelima Dubey. B: *The Journal of Chemical Thermodynamics* 40 (2008), с. 309—320. DOI: <https://doi.org/10.1016/j.jct.2007.05.016>.
- [18] Gyan P. Dubey и Monika Sharma. B: *The Journal of Chemical Thermodynamics* 40 (2008), с. 991—1000. DOI: <https://doi.org/10.1016/j.jct.2008.02.005>.
- [19] K.V.N. Suresh Reddy, G. Sankara Reddy и A. Krishnaiah. B: *Thermochimica Acta* 440 (2006), с. 43—50. DOI: <https://doi.org/10.1016/j.tca.2005.10.008>.
- [20] M. Domínguez и др. B: *J. Chem. Eng. Data* 48 (2003), с. 302—307. DOI: 10.1021/je020114l.
- [21] A. Rodríguez и др. B: *J. Chem. Eng. Data* 48 (2003), с. 146—151. DOI: 10.1021/je020131a.
- [22] Jyoti N. Nayak, Mrityunjaya I. Aralaguppi и Tejraj M. Aminabhavi. B: *J. Chem. Eng. Data* 48 (2003), с. 1152—1156. DOI: 10.1021/je030107c.
- [23] Mikhail F. Bolotnikov и Yuriy A. Neruchev. B: *J. Chem. Eng. Data* 48 (2003), с. 739—741. DOI: 10.1021/je034002l.
- [24] Ignacio Gascón и др. B: *J. Chem. Eng. Data* 50 (2005), с. 722—726. DOI: 10.1021/je049576k.
- [25] Changsheng Yang, Wei Xu и Peisheng Ma. B: *J. Chem. Eng. Data* 49 (2004), с. 1802—1808. DOI: 10.1021/je049777o.
- [26] Gyan P. Dubey и Monika Sharma. B: *J. Chem. Eng. Data* 52 (2007), с. 449—453. DOI: 10.1021/je060389r.
- [27] Qinglai Tian и Huizhou Liu. B: *J. Chem. Eng. Data* 52 (2007), с. 892—897. DOI: 10.1021/je060491o.
- [28] Ana Carrasco и др. B: *J. Chem. Eng. Data* 53 (2008), с. 1223—1227. DOI: 10.1021/je800048f.
- [29] Sheng Fang и др. B: *J. Chem. Eng. Data* 53 (2008), с. 2718—2720. DOI: 10.1021/je8006138.
- [30] Manapragada V. Rathnam, Sharad Mankumare и M. S. S. Kumar. B: *J. Chem. Eng. Data* 55 (2010), с. 1354—1358. DOI: 10.1021/je9006597.
- [31] N. Tripathi. «Densities, Viscosities, and Refractive Indices of Mixtures of Hexane with Cyclohexane, Decane, Hexadecane, and Squalane at 298.15 K». B: *International Journal of Thermophysics* 26 (2005), с. 693—703. DOI: 10.1007/s10765-005-5572-8.
- [32] Montserrat Domínguez-Pérez и др. «Experimental Study of the Dynamic Viscosity Deviations in the Binary Systems: Hexane + Ethylbenzene, + o-Xylene, + m-Xylene, + p-Xylene at 298.15 K». B: *International Journal of Thermophysics* 30 (2009), с. 1197—1201. DOI: 10.1007/s10765-009-0622-2.
- [33] Fátima Reyes-García и Gustavo A. Iglesias-Silva. B: *J. Chem. Eng. Data* 62 (2017), с. 2726—2739. DOI: 10.1021/acs.jced.7b00121.
- [34] Claus K. Zéberg-Mikkelsen и др. B: *Fluid Phase Equilibria* 245 (2006), с. 6—19. DOI: <https://doi.org/10.1016/j.fluid.2006.01.030>.
- [35] Emila M. Živković и др. B: *Fluid Phase Equilibria* 299 (2010), с. 191—197. DOI: <https://doi.org/10.1016/j.fluid.2010.10.009>.

- [36] Wanchalerm Srirachat и др. В: *Fluid Phase Equilibria* 434 (2017), с. 117—129. DOI: <https://doi.org/10.1016/j.fluid.2016.11.029>.
- [37] I.M. Abdulagatov и N.D. Azizov. В: *The Journal of Chemical Thermodynamics* 38 (2006), с. 1402—1415. DOI: <https://doi.org/10.1016/j.jct.2006.01.012>.
- [38] Sheng Fang и др. В: *The Journal of Chemical Thermodynamics* 68 (2014), с. 281—287. DOI: <https://doi.org/10.1016/j.jct.2013.09.017>.
- [39] Johnny R. Zambrano и др. В: *The Journal of Chemical Thermodynamics* 96 (2016), с. 104—116. DOI: <https://doi.org/10.1016/j.jct.2015.12.021>.
- [40] Zhuqiang Yang и др. В: *Thermochimica Acta* 617 (2015), с. 1—7. DOI: <https://doi.org/10.1016/j.tca.2015.08.005>.
- [41] Chao Su и др. В: *Thermochimica Acta* 670 (2018), с. 211—218. DOI: <https://doi.org/10.1016/j.tca.2018.10.018>.
- [42] Alfonso S. Pensado и др. В: *J. Chem. Eng. Data* 50 (2005), с. 849—855. DOI: 10.1021/je049662k.
- [43] Changsheng Yang, Wei Xu и Peisheng Ma. В: *J. Chem. Eng. Data* 49 (2004), с. 1794—1801. DOI: 10.1021/je049776w.
- [44] Hong Liu и Lin Zhu. В: *J. Chem. Eng. Data* 59 (2014), с. 369—375. DOI: 10.1021/je400835u.
- [45] Dianne J. Luning Prak, Jim S. Cowart и Paul C. Trulove. В: *J. Chem. Eng. Data* 59 (2014), с. 3842—3851. DOI: 10.1021/je5007532.
- [46] Diana C. Landaverde-Cortes и др. В: *J. Chem. Eng. Data* 52 (2007), с. 1226—1232. DOI: 10.1021/je600554h.
- [47] Zhi-Fei Wang, Li-Sheng Wang и Tian-Bo Fan. В: *J. Chem. Eng. Data* 52 (2007), с. 1866—1871. DOI: 10.1021/je700202h.
- [48] Sheng Fang, Chun-Xia Zhao и Chao-Hong He. В: *J. Chem. Eng. Data* 53 (2008), с. 2244—2246. DOI: 10.1021/je8003707.
- [49] Xi Yang и др. В: *J. Chem. Eng. Data* 55 (2010), с. 2914—2916. DOI: 10.1021/je900969u.
- [50] Hanan E. M. El-Sayed и Abdul-Fattah A. Asfour. «Viscometric and Volumetric Properties of 10 Regular Binary Systems at 308.15 K and 313.15 K». В: *International Journal of Thermophysics* 30 (2009), с. 1773—1790. DOI: 10.1007/s10765-009-0667-2.
- [51] D. I. Sagdeev и др. «Experimental Study of the Density and Viscosity of

*n*

- Heptane at Temperatures from 298 K to 470 K and Pressure upto 245 MPa». В: *International Journal of Thermophysics* 34 (2013), с. 1—33. DOI: 10.1007/s10765-012-1373-z.
- [52] Omar El-Hadad, Ruo Cai и Abdul-Fattah A. Asfour. «Densities and Kinematic Viscosities of a Quinary Regular Liquid System and Its Five Quaternary Subsystems at 293.15 K and 298.15 K». В: *International Journal of Thermophysics* 36 (2015), с. 69—80. DOI: 10.1007/s10765-014-1759-1.
- [53] Zhuqiang Yang, Qincheng Bi и Song Feng. В: *J. Chem. Eng. Data* 61 (2016), с. 3472—3480. DOI: 10.1021/acs.jced.6b00391.

- [54] B. González, A. Domínguez и J. Tojo. В: *The Journal of Chemical Thermodynamics* 36 (2004), с. 267—275. DOI: <https://doi.org/10.1016/j.jct.2003.12.005>.
- [55] Lei Yue и др. В: *The Journal of Chemical Thermodynamics* 81 (2015), с. 26—33. DOI: <https://doi.org/10.1016/j.jct.2014.09.015>.
- [56] José L. Trenzado и др. В: *J. Chem. Eng. Data* 48 (2003), с. 1004—1014. DOI: 10.1021/je034017j.
- [57] Begoña González и др. В: *J. Chem. Eng. Data* 49 (2004), с. 1225—1230. DOI: 10.1021/je034208m.
- [58] Changsheng Yang, Wali Yu и Peisheng Ma. В: *J. Chem. Eng. Data* 50 (2005), с. 1197—1203. DOI: 10.1021/je049572f.
- [59] Anthony R. H. Goodwin и др. В: *J. Chem. Eng. Data* 51 (2006), с. 190—208. DOI: 10.1021/je0503296.
- [60] Xianjie Gong и др. В: *J. Chem. Eng. Data* 57 (2012), с. 3278—3282. DOI: 10.1021/je300899n.
- [61] A. Estrada-Baltazar, G. A. Iglesias-Silva и C. Caballero-Cerón. В: *J. Chem. Eng. Data* 58 (2013), с. 3351—3363. DOI: 10.1021/je4004806.
- [62] Fengjun Yang и др. В: *J. Chem. Eng. Data* 53 (2008), с. 2237—2240. DOI: 10.1021/je800348s.
- [63] Derek R. Caudwell и др. В: *J. Chem. Eng. Data* 54 (2009), с. 359—366. DOI: 10.1021/je800417q.
- [64] Yoshiyuki Sato и др. «A Digital Variable-Angle Rolling-Ball Viscometer for Measurement of Viscosity, Density, and Bubble-Point Pressure of CO<sub>2</sub> and Organic Liquid Mixtures». В: *International Journal of Thermophysics* 31 (2010), с. 1896—1903. DOI: 10.1007/s10765-008-0542-6.
- [65] Lingling Zhang и др. В: *J. Chem. Eng. Data* 56 (2011), с. 4268—4273. DOI: 10.1021/je200757a.
- [66] Hai Chi и др. В: *J. Chem. Eng. Data* 58 (2013), с. 2224—2232. DOI: 10.1021/je400250u.
- [67] Diana C. Landaverde-Cortes и др. В: *J. Chem. Eng. Data* 53 (2008), с. 288—292. DOI: 10.1021/je700428f.
- [68] Zhaohui Liu, J. P. Martin Trusler и Qincheng Bi. В: *J. Chem. Eng. Data* 60 (2015), с. 2363—2370. DOI: 10.1021/acs.jced.5b00270.
- [69] Dianne J. Luning Prak и др. В: *J. Chem. Eng. Data* 62 (2017), с. 169—187. DOI: 10.1021/acs.jced.6b00542.
- [70] Aravind R. Mahajan и Sunil R. Mirgane. В: *The Journal of Chemical Thermodynamics* 64 (2013), с. 159—166. DOI: <https://doi.org/10.1016/j.jct.2013.05.014>.
- [71] Seema Kapoor и Virender K. Rattan. В: *J. Chem. Eng. Data* 50 (2005), с. 1891—1896. DOI: 10.1021/je0501585.
- [72] Xiaomei Qin и др. В: *J. Chem. Eng. Data* 59 (2014), с. 775—783. DOI: 10.1021/je4008926.
- [73] Zhiping Fang и др. В: *J. Chem. Eng. Data* 53 (2008), с. 2787—2792. DOI: 10.1021/je800635g.

- [74] Jing Zhao и др. В: *J. Chem. Eng. Data* 62 (2017), с. 643—652. DOI: 10.1021/acs.jced.6b00688.
- [75] Xinxin Cheng и др. В: *J. Chem. Eng. Data* 62 (2017), с. 2330—2339. DOI: 10.1021/acs.jced.7b00201.
- [76] Luning и др. В: *J. Chem. Eng. Data* 62 (2017), с. 3452—3472. DOI: 10.1021/acs.jced.7b00466.
- [77] Song Feng и др. В: *J. Chem. Eng. Data* 63 (2018), с. 671—678. DOI: 10.1021/acs.jced.7b00866.
- [78] Dianne J. Luning Prak и др. В: *J. Chem. Eng. Data* 63 (2018), с. 1642—1656. DOI: 10.1021/acs.jced.8b00008.
- [79] Yitong Dai и др. В: *J. Chem. Eng. Data* 63 (2018), с. 4052—4060. DOI: 10.1021/acs.jced.8b00438.
- [80] Dianne J. Luning Prak и др. В: *J. Chem. Eng. Data* 64 (2019), с. 1550—1568. DOI: 10.1021/acs.jced.8b01135.
- [81] Dianne J. Luning Prak и др. В: *J. Chem. Eng. Data* 64 (2019), с. 1725—1745. DOI: 10.1021/acs.jced.8b01233.
- [82] Fernando Czubinski и др. В: *J. Chem. Eng. Data* 64 (2019), с. 3375—3384. DOI: 10.1021/acs.jced.9b00187.
- [83] Iria Rodríguez-Escontrela и др. В: *Fluid Phase Equilibria* 405 (2015), с. 124—131. DOI: <https://doi.org/10.1016/j.fluid.2015.07.022>.
- [84] Marek Blahušiak и Štefan Schlosser. В: *The Journal of Chemical Thermodynamics* 72 (2014), с. 54—64. DOI: <https://doi.org/10.1016/j.jct.2013.12.022>.
- [85] Sara Lago и др. В: *The Journal of Chemical Thermodynamics* 75 (2014), с. 63—68. DOI: <https://doi.org/10.1016/j.jct.2014.02.012>.
- [86] Mani Lal Singh и др. В: *J. Chem. Eng. Data* 59 (2014), с. 1130—1139. DOI: 10.1021/je400493x.
- [87] Dianne J. Luning Prak и др. В: *J. Chem. Eng. Data* 59 (2014), с. 1334—1346. DOI: 10.1021/je5000132.