



НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ
УНИВЕРСИТЕТ



NORTH-EASTERN FEDERAL
UNIVERSITY IN YAKUTSK

Северо-Восточный федеральный университет Институт математики и информатики

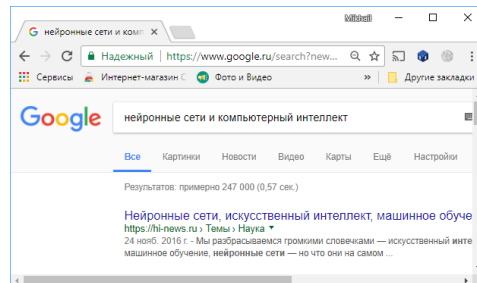
Михаил Антонов
научный сотрудник

e-mail: mikhail@s-vfu.ru
T: +7 (4112) 49-68-33



Краткая история

- Зародившись в 1960-х годах как анализ данных, «Наука о данных» и «Машинное обучение» получили широкую популярность в 2010-х годах.
- McKinsey & Company оценивала потребность в специалистах в области Big Data в 1,5 млн человек только в США в 2018 г.



Образовательные программы

- С 2013 года специализированные образовательные программы открыты ведущих западных учебных заведениях, сегодня имеются и в ведущих вузах РФ – МГУ, ВШЭ, МФТИ, ИТМО и др.



Stanford
University



Berkeley
UNIVERSITY OF CALIFORNIA

Основные идеи алгоритмов МО

Что могут алгоритмы МО.

Прямые задачи:

За сколько времени Саша придет из пункта А в пункт Б со скоростью x_1 , если расстояние между А и Б x_2

$$F: X(x_1, x_2) \rightarrow Y(y)$$



Основные идеи алгоритмов МО

Что могут алгоритмы МО.

Прямые задачи:

За сколько времени Саша придет из пункта А в пункт Б со скоростью x_1 , если расстояние между А и Б x_2

$F: X(x_1, x_2) \rightarrow Y(y)$

Решение: $y = x_2 / x_1$



Основные идеи алгоритмов МО

Что могут алгоритмы МО.

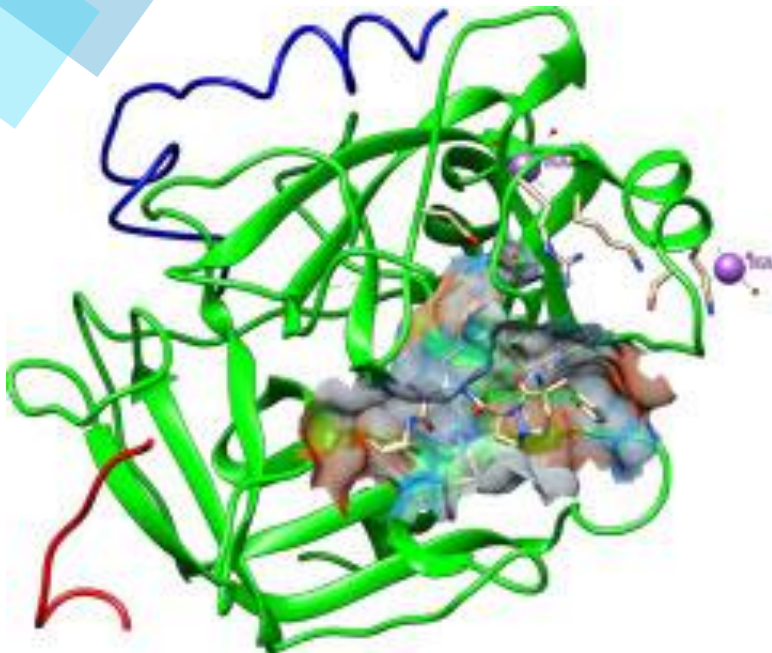
$$\vec{F}_i(\vec{r}) = -\frac{\partial U(\vec{r})}{\partial \vec{r}_i} + \vec{F}_{ext}$$

$$m_i \frac{d^2 \vec{r}_i}{dt^2} = \vec{F}_i$$

$$U = U_b + U_v + U_\phi + U_{vw} + U_{qq} + U_f$$

$$U_\phi = \frac{1}{2} \sum_{\phi=1}^{\Phi} \sum_{l=1}^L K_{\phi,l} \left[1 + g_{\phi,l} \cos(n_{\phi,l} \alpha) \right]$$

$$U_{vw} = \sum_{i \neq j} \left[\frac{A}{r_{i,j}^{12}} - \frac{B}{r_{i,j}^6} \right] \quad U_v = \frac{1}{2} \sum_{v=1}^V K_v (\alpha - \alpha_0)^2 \quad U_{qq} = \sum_{i \neq j} \frac{q_i q_j}{\epsilon r_{i,j}}$$



Основные идеи алгоритмов МО

Что могут алгоритмы МО.

Представим, что параметров огромное количество.

Мы не знаем связи между параметрами и результатом.



Модели машинного обучения

- Провести миллион миллиардов экспериментов (или у нас есть накопленные результаты)
- Обработать результаты компьютером
- Автоматически найти все зависимости
- Получить ответы на новых данных



Определение ML

Машинное обучение — класс методов искусственного интеллекта (ML), характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач.



ВИКИПЕДИЯ
Свободная энциклопедия

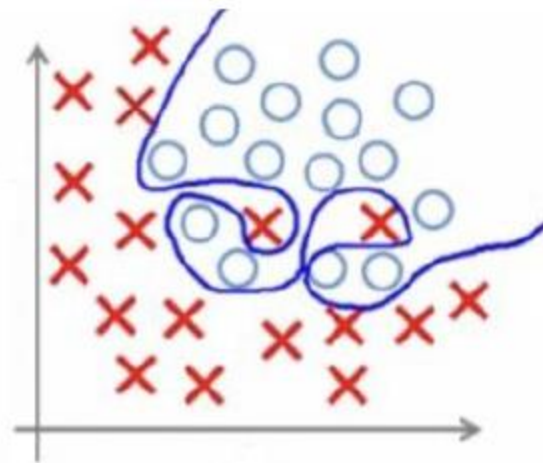
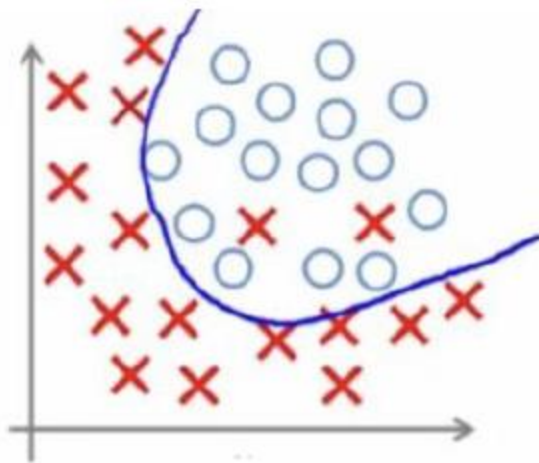
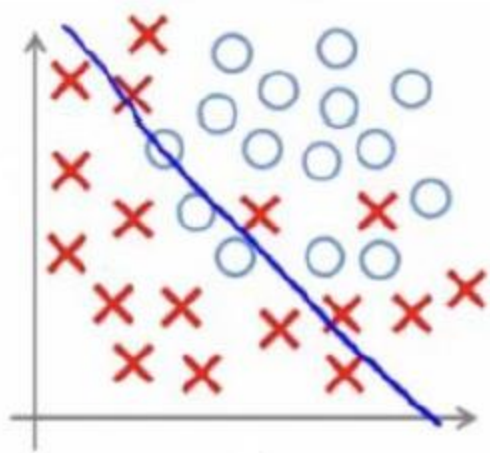
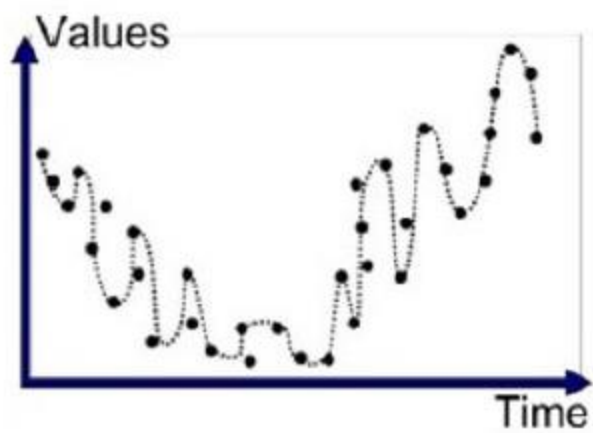
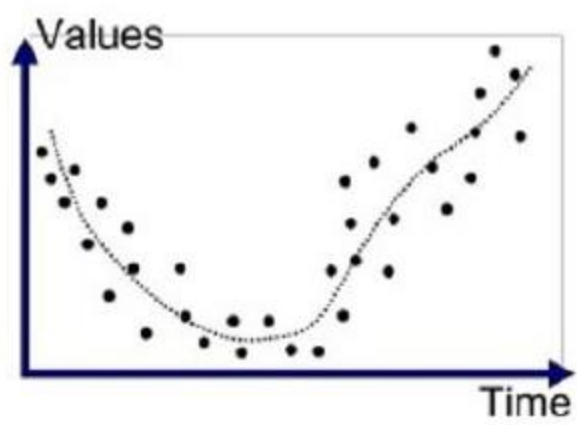
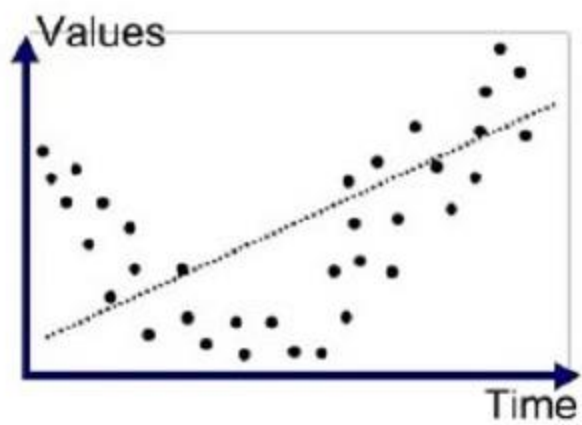
https://ru.wikipedia.org/wiki/Машинное_обучение

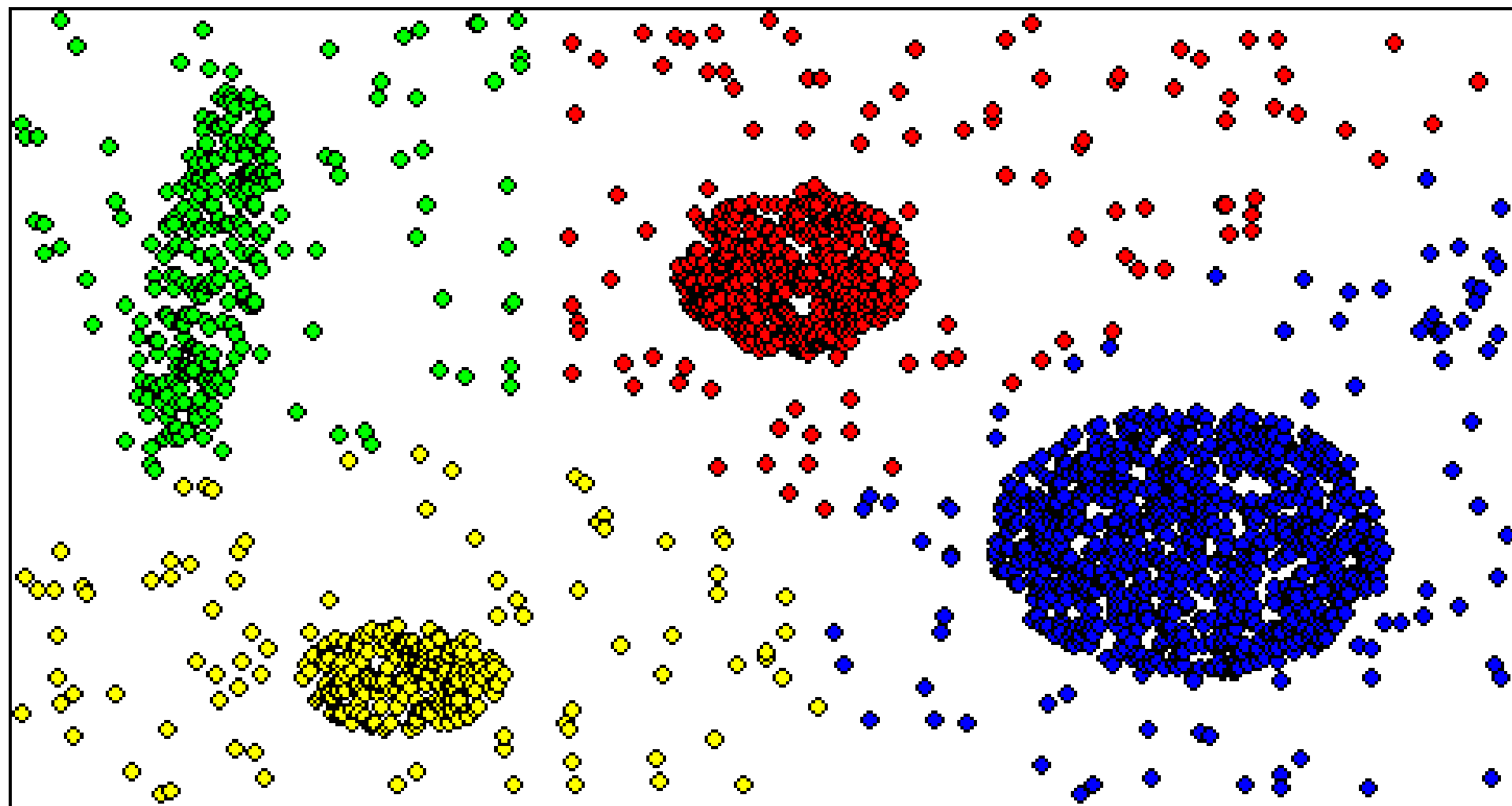


Определение ML

Задача: восстановить сложную зависимость по конечному числу примеров.

- Задачи классификации
- Задачи регрессии
- Задачи кластеризации
- Задачи восстановления данных
- Задачи поиска выбросов
- ...





Образовательные программы



Образовательная программа Наука о данных и машинное обучение



Шамаев Эллей Иванович

Академический руководитель
образовательной программы



datascience_nefu

Образовательная программа

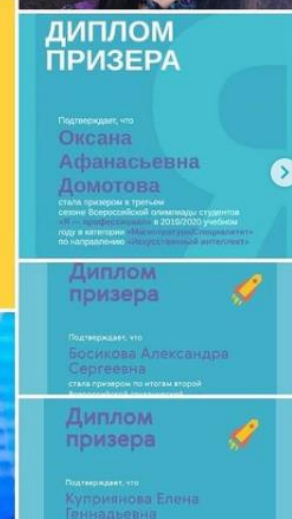
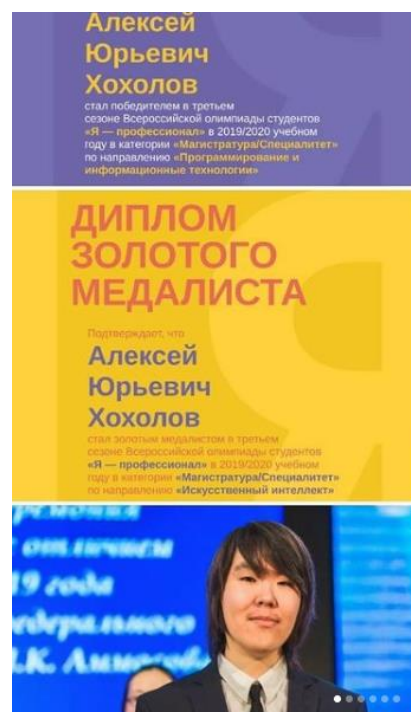
Наука о данных и машинное обучение



 **datascience_nefu** Поздравляем студента математического отделения ИМИ СВФУ Ньургун Федорова! Его работа "Сверточные нейросети в сортировке алмазного сырья по качеству" оценена дипломом третьей степени на престижной студенческой конференции МНСК-2020 Новосибирского госуниверситета.

Ньургун Федоров вовлечен в совместные проекты с нашими магистрантами. Диплом МНСК вручен за исследования после реального проекта в конкурсе АЛРОСА летом 2019 года.

   25 отзывов "Ньюс" [



Образовательные программы IT Samsung Academy



- Академия Мобильной разработки
- Академия Интернета Вещей
- Академия Искусственного интеллекта

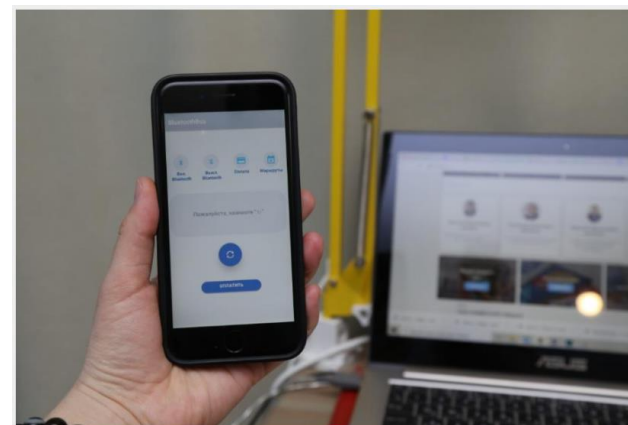


фото: Нина СЛЕПЦОВА / редакция корпоративных медиа СВФУ

3 ноября 2020 449

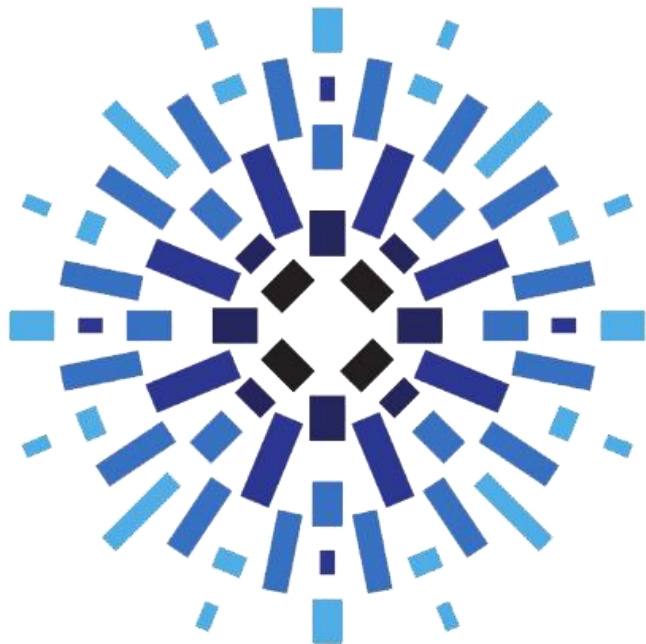
Команда СВФУ прошла в финал конкурса IT Академии Samsung



Образовательные программы Искусственный интеллект АСИ

20.35
УНИВЕРСИТЕТ

Консорциум университетов Исследователей больших данных

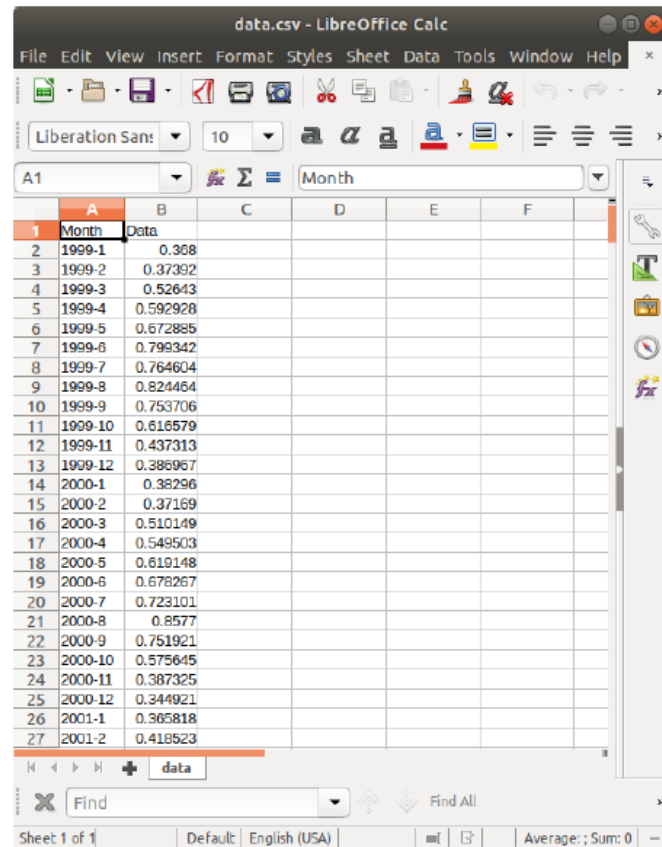


Университетский
консорциум
исследователей
больших данных



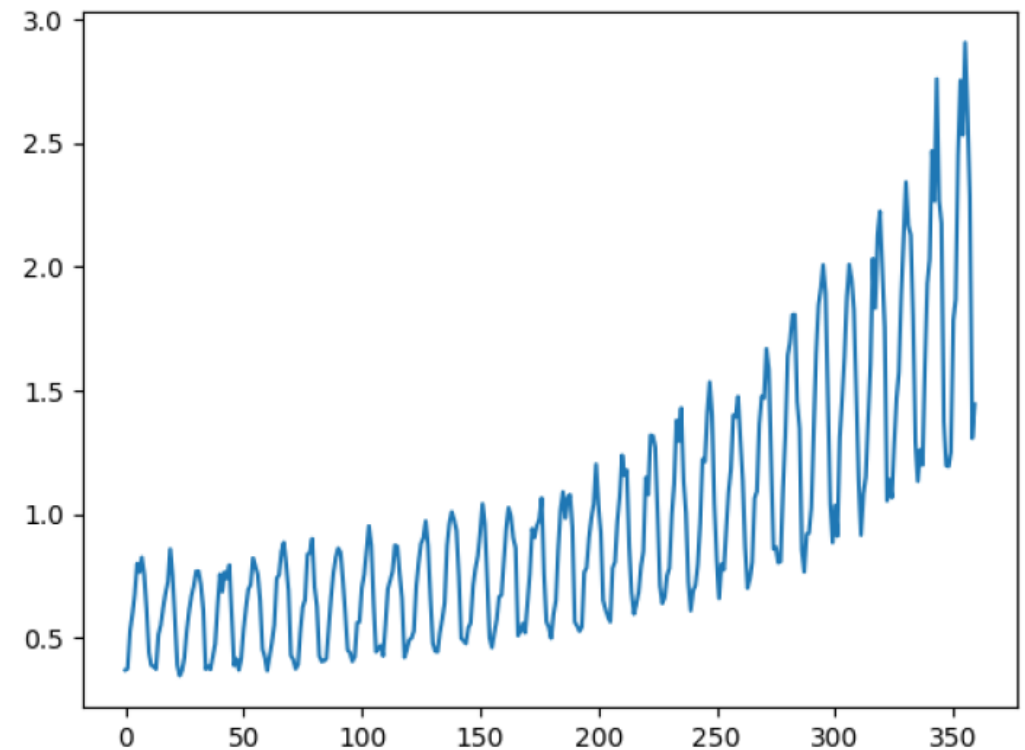
Научные исследования

Предсказание последовательностей с помощью рекуррентных нейронных сетей



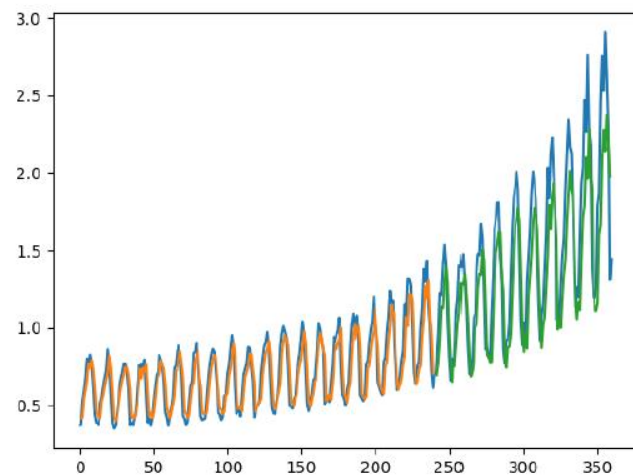
Month	Data
1999-1	0.368
1999-2	0.37392
1999-3	0.52643
1999-4	0.592928
1999-5	0.672885
1999-6	0.799342
1999-7	0.764604
1999-8	0.824464
1999-9	0.753706
1999-10	0.616579
1999-11	0.437313
1999-12	0.385957
2000-1	0.38296
2000-2	0.37169
2000-3	0.510149
2000-4	0.549503
2000-5	0.619148
2000-6	0.678267
2000-7	0.723101
2000-8	0.8577
2000-9	0.751921
2000-10	0.575645
2000-11	0.387325
2000-12	0.344921
2001-1	0.365818
2001-2	0.418523

(a) $view = 1$

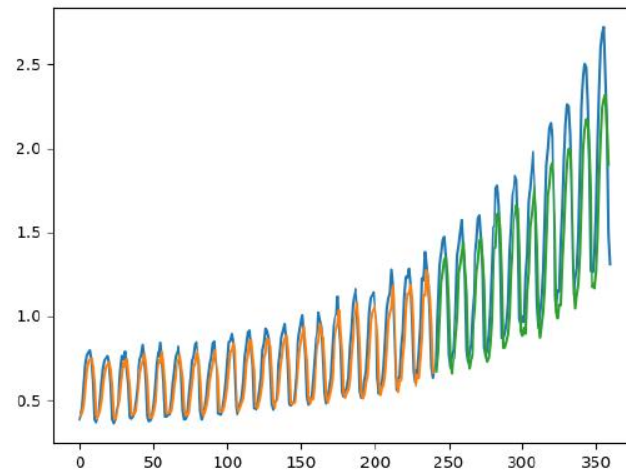


(b) $view = 2$

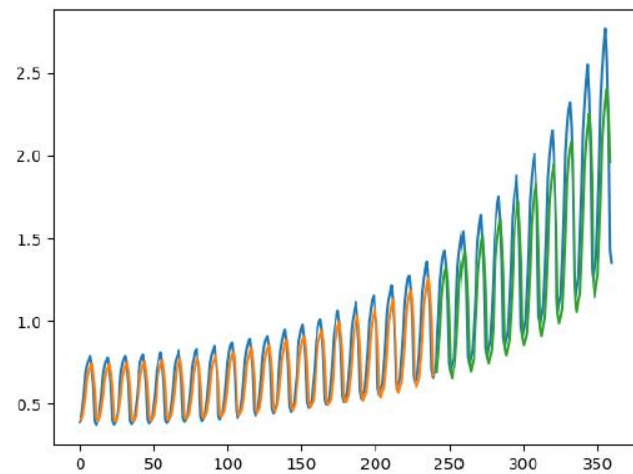
Научные исследования



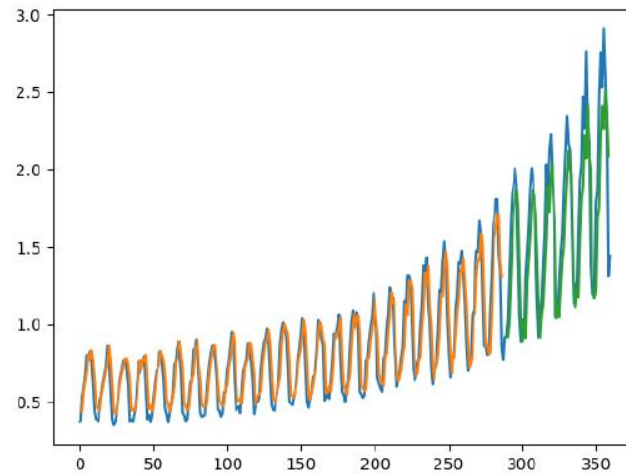
Noise 10%



Noise 5%



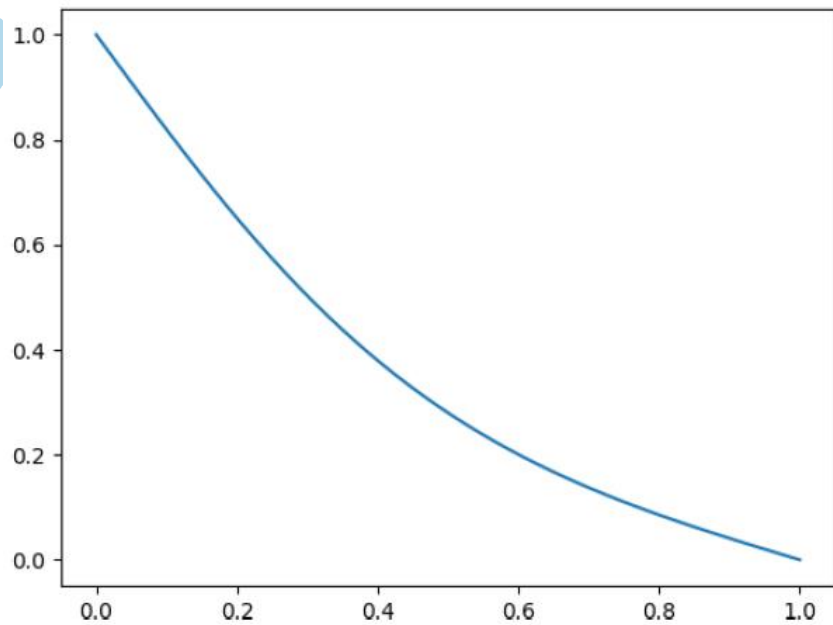
Noise 1%



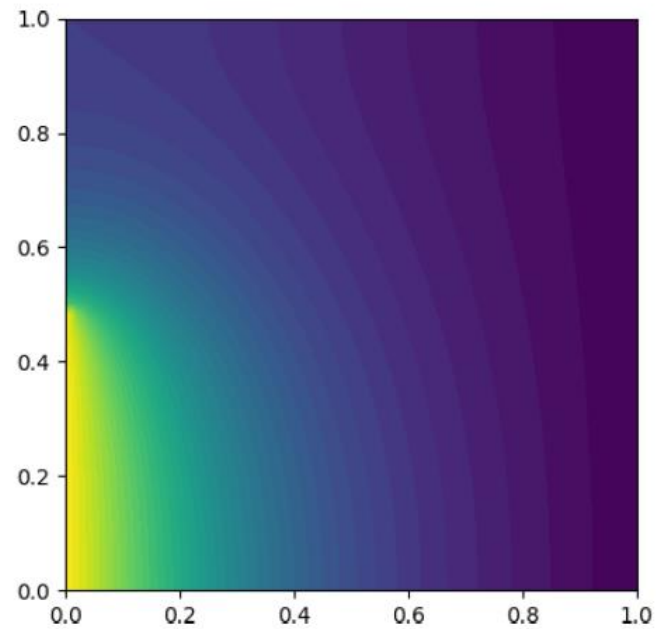
Noise 10% (Training 80%)

Научные исследования

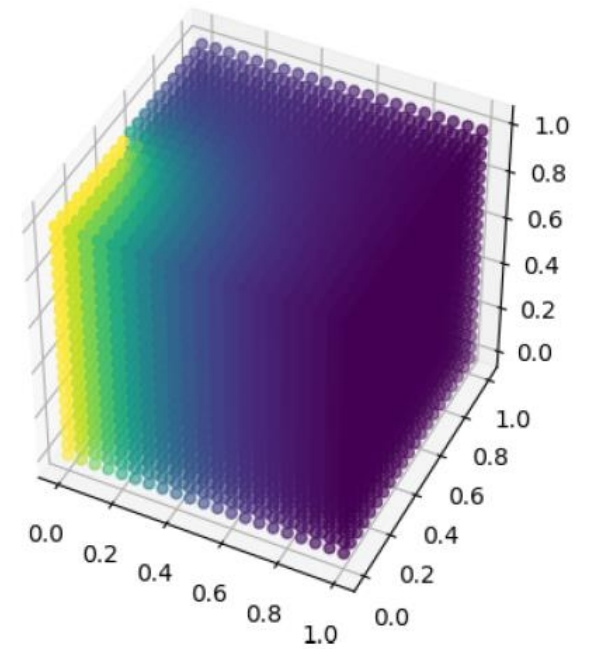
Решение нелинейной задачи течения при гидроразрыве



1D



2D



3D

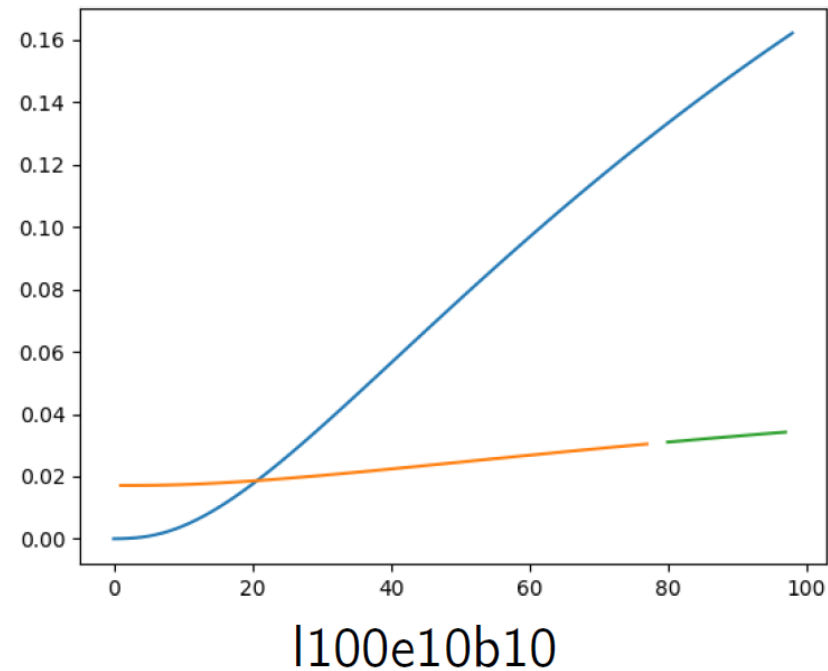
Научные исследования

Решение нелинейной задачи течения при гидроразрыве

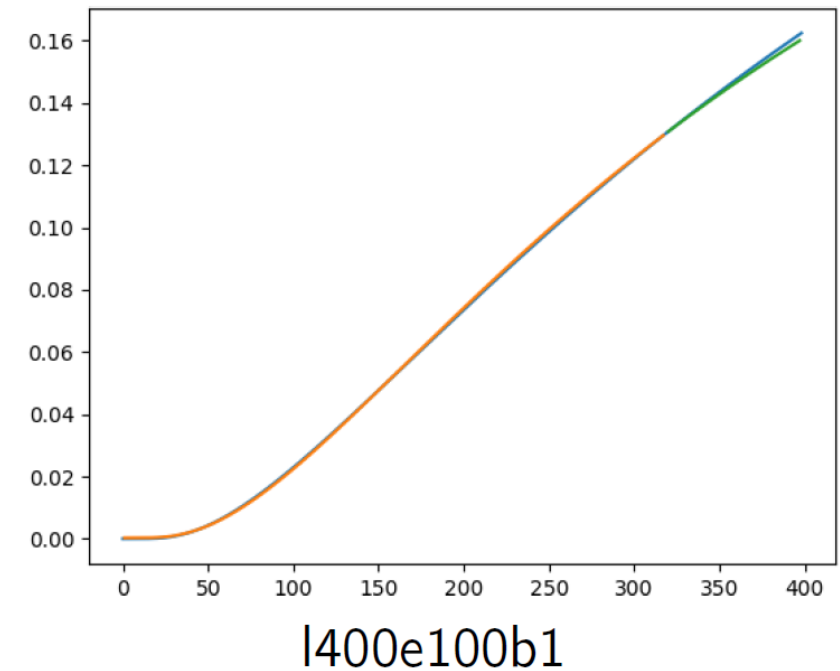
Университет Сянтань (КНР)



bad accuracy
loss = 0.11 RMSE



good accuracy
loss = 10E-5 RMSE



Научные исследования

Международный грант совместно с коллективом из Вьетнама РФФИ, 21-51-54001 Вьет_а, Новые математические модели для многомасштабного процесса просачивания с применением подходов машинного обучения.

Institute of Research and Development, Duy Tan University, Room 809, 03 Quang Trung, Da Nang 550000, Vietnam

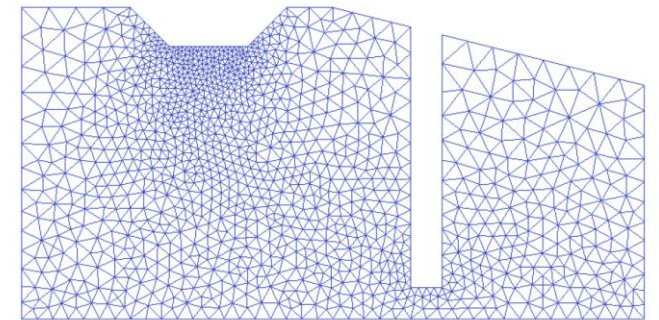
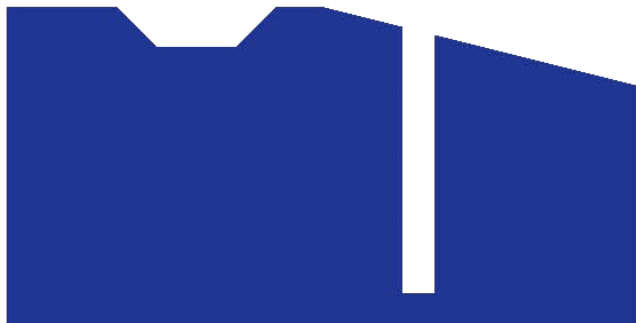


Figure: Расчетная сетка: 1292 узлов, 2430 элемента

Исследовательская задача

- Целевые переменные: theses и prsactice
- Объекты соответствуют вектору студента.
- 'theses', 'practice' - целевые переменные: Оценка за диссертацию и преддипломную практику
- 'problems_number_i' - Количество "долгов" перед сессией
- 'subject_number_i' - Количество дисциплин
- 'mark_10_average_i' - средняя оценка
- 'try_number_i' - количество попыток пересдачи
- i - номер семестра

Исследовательская задача



CP.3H (th) = 8.59

CP.3H (pr) = 8.92

CP.ОТКЛ (th) = 1.08

CP.ОТКР (pr) = 0.95

N=77

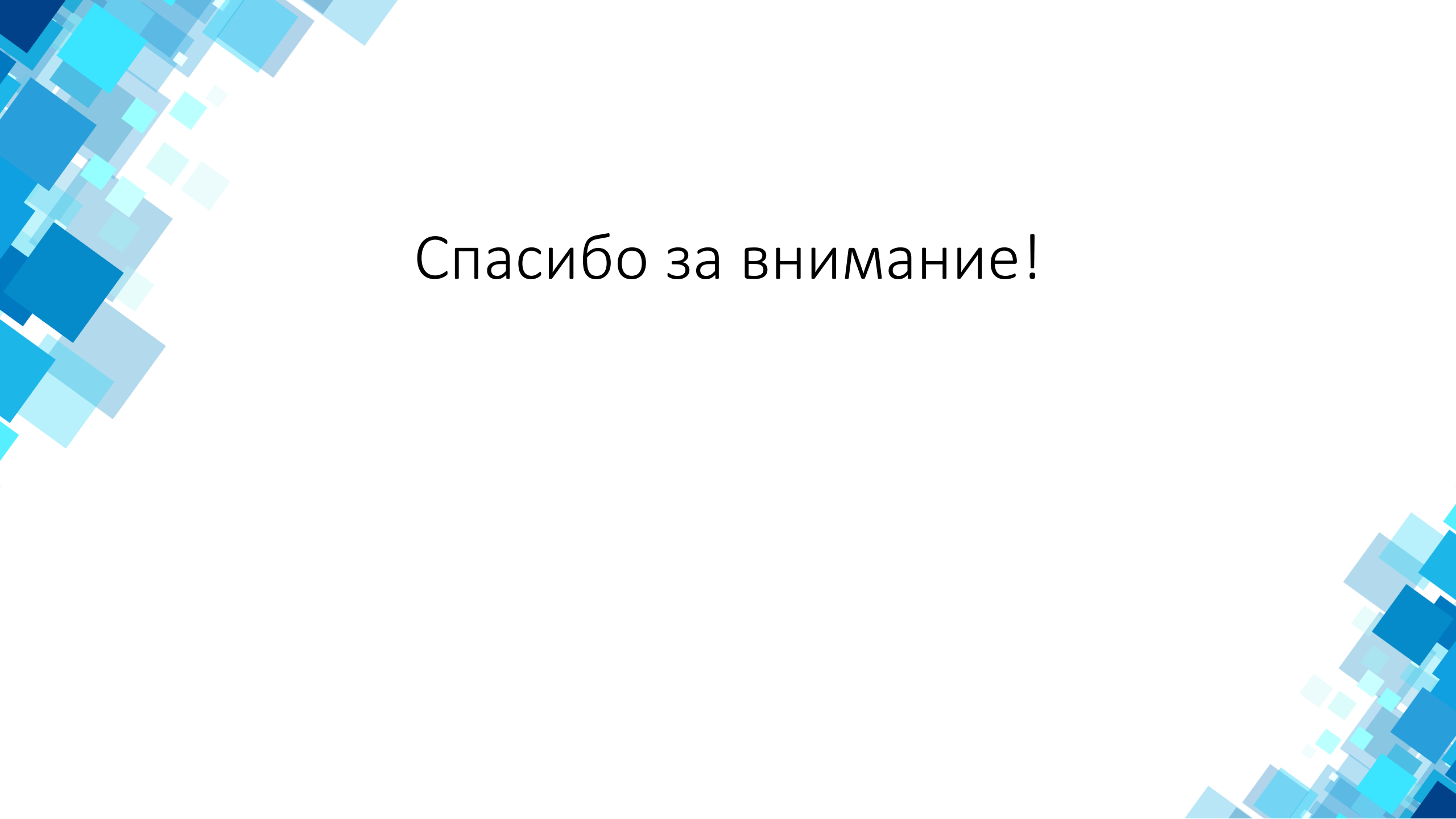
	A	AJ	AK	AL	AM	AN	AO	AP	AQ	AR	AS	AT	AU
1	student_id	mark_10	try_number	problems	subject_number	mark_10	try_number	problems	subject_number	mark_10	try_number	theses	practice
2	id541004											9	10
3	id220206	8.8	0	0	8	9	0	0	3	9.33333	0	9	10
4	id622235	8.8	0	0	8	9	0	0	3	9.33333	0	9	10
5	id125876	6.75	0	0	13	6.92308	0	0	5	8	0	6	6
6	id829539	8.375	0	0	11	7.63636	0	0	3	9.33333	0	8	10
7	id122434	6.75	0	1	13	4.69231	1	0	6	8	0	3	9
8	id551371	8.14286	0	0	10	8	0	0	6	7.66667	0	10	10
9	id806077	8.6	0	0	11	8.45455	0	0	5	8.2	0	8	9
10	id539778	8.6	0	0	11	8.45455	0	0	5	8.2	0	8	9
11	id468131	7.66667	0	0	10	6.7	1	0	3	7.66667	0	9	8
12	id536330	6.85714	0	0	11	7.54545	0	0	3	8.66667	0	10	10
13	id661127	6.85714	0	0	11	7.54545	0	0	3	8.66667	0	10	10
14	id852154	8.28571	0	0	10	9	0	0	6	9.16667	0	10	10
15	id770414	7.8	0	0	8	8.375	0	0	3	8.33333	0	10	10
16	id550507	8.33333	0	0	10	8.3	0	0	6	9.5	0	10	10
17	id105876	6	0	1	12	4.16667	1	0	6	7.5	0	8	8
18	id218804	8.16667	0	0	10	8.2	0	0	6	8.83333	0	8	9
19	id703700	9.28571	0	0	11	9.18182	0	0	3	9.66667	0	10	10
20	id463279	9.28571	0	0	11	9.18182	0	0	3	9.66667	0	10	10
21	id795934	8.5	0	0	11	9	0	0	3	8.66667	0	10	9
22	id487534	7.375	0	0	11	7.09091	0	0	3	8.33333	0	10	9
23	id474886	7.375	0	0	11	7.09091	0	0	3	8.33333	0	10	9
24	id482580	7.375	0	0	11	7.09091	0	0	3	8.33333	0	10	9
25	id610700	8.42857	0	0	11	8.45455	0	0	3	8.66667	0	8	10
26	id870385	7.5	0	0	8	8.125	0	0	3	8.66667	0	6	7

kNN, 5-fold crossval

k	MAE th (p=1)	MAE pr (p=1)	MAE th (p=2)	MAE pr (p=2)	k	MAE th (p=1)	MAE pr (p=1)	MAE th (p=2)	MAE pr (p=2)
2	-1.1	-1.0	-1.2	-0.8	9	-0.644	-0.64	-0.77	-0.64
3	-1.033	-0.73	-1.13	-0.66	10	-0.61	-0.7	-0.77	-0.68
4	-0.775	-0.7	-0.92	-0.55	11	-0.64	-0.69	-0.75	-0.62
5	-0.78	-0.68	-0.82	-0.7	12	-0.62	-0.62	-0.74	-0.65
6	-0.766	-0.68	-0.81	-0.7	13	-0.7	-0.59	-0.69	-0.62
7	-0.742	-0.68	-0.74	-0.65	14	-0.73	-0.63	-0.74	-0.61
8	-0.75	-0.65	-0.66	-0.60	15	-0.75	-0.65	-0.72	-0.58

RF, 5-fold crossval

Ntr	MAE th	MAE pr	Ntr	MAE th	MAE pr
5	-0.68 (Max.Depth. 10)	-0.67 (10)	35	-0.66 (19)	-0.78 (16)
7	-0.65 (Max.Depth. 19)	-0.75 (11)	55	-0.66 (12)	-0.81 (11)
9	-0.71 (Max.Depth. 14)	-0.58 (18)	85	-0.65 (14)	-0.80 (15)
11	-0.73 (Max.Depth. 17)	-0.76 (10)	125	-0.69 (18)	-0.78 (18)
13	-0.59 (Max.Depth. 10)	-0.73 (18)	155	-0.69 (17)	-0.83 (11)
15	-0.61 (Max.Depth. 13)	-0.72 (10)	175	-0.70 (13)	-0.83 (18)
25	-0.71 (Max.Depth. 14)	-0.78 (10)	199	-0.71 (17)	-0.84 (11)

The image features a light blue background with decorative elements in the corners. The top-left and bottom-right corners are filled with a cluster of overlapping squares in various shades of blue, ranging from light cyan to dark navy. The text is centered in the middle of the slide.

Спасибо за внимание!