

УДК 519.25

**ТЕХНОЛОГИИ БОЛЬШИХ ДАННЫХ ДЛЯ АНАЛИЗА ВЛИЯНИЯ
КОМПОНЕНТОВ ОБРАЗОВАТЕЛЬНОЙ СРЕДЫ НА
РЕЗУЛЬТАТИВНОСТЬ ПРОФЕССИОНАЛЬНОЙ ДЕЯТЕЛЬНОСТИ
УЧИТЕЛЕЙ¹⁾**

А.И. МУЗАФАРОВА¹⁾, Ф.М. ГАФАРОВ¹⁾

¹⁾ *Казанский (Приволжский) федеральный университет, 420008, г. Казань, ул.
Кремлевская, 18,
E-mail: Alina-291997muzafarova@yandex.ru*

**BIG DATA TECHNOLOGIES FOR ANALYZING THE IMPACT OF THE
COMPONENTS OF THE EDUCATIONAL ENVIRONMENT ON THE
PERFORMANCE OF TEACHERS' PROFESSIONAL ACTIVITIES**

A.I. MUZAFAROVA¹⁾, F.M. GAFAROV¹⁾

¹⁾ *Kazan Federal University, 18 Kremlyovskaya Street, 420008, Kazan, Russia*

Аннотация

Статья посвящена применению технологий "большие данные" (Big Data) для анализа данных в сфере школьного образования. Подробно описана информационно-аналитическая система, построенная для проведения интеллектуального анализа данных об образовательном процессе, на основе массивов больших данных о деятельности педагога, администрации школы и успешности учащихся. Анализ проведен на основе данных собранных в системе "Электронное образование для Республики Татарстан" за период с 2015 по 2020 годы. Приводятся алгоритмы обработки данных, на основе которых можно анализировать и прогнозировать сложившуюся обстановку в современном образовательном пространстве. Описываются возможности высокоуровневой библиотеки Pandas для обработки и анализа данных и гибкой библиотеки Dask для проведения параллельных вычислений в кластерных вычислительных системах. Проведен корреляционный анализ факторов влияющих на эффективность профессиональной деятельности учителя и успешность обучения школьников на основе коэффициентов корреляции Пирсона, Спирмена и Кендалла. Для проверки гипотезы о равенстве генеральных средних двух независимых выборок к данным был применен t-критерий Стьюдента. Выявлены и исследованы закономерности, которым подчиняются процессы в системе образования.

Ключевые слова: Большие данные, Dask, Pandas, образовательный процесс, корреляционный анализ, профессиональная успешность учителя

Summary

The article is devoted to the application of Big Data technologies for data analysis in the field of school education. The information- analytical system, built for the intellectual analysis of data on educational processes, based on the big data arrays on the activities of the teacher, the school administration and the success of the students, is described in detail. The analysis was carried out on the basis of data collected in the "Electronic Education for the Republic of Tatarstan" system for the period from 2015 to 2020. Algorithms for data processing are given, on the basis of which it is possible to analyze and predict the current situation in the modern educational space. It describes the capabilities of the high-level Pandas library for data processing and analysis and the flexible Dask

¹⁾ Исследование выполнено при финансовой поддержке РФФИ в рамках научного проекта № 19-29-14082

library for parallel computing in cluster computing systems. A correlation analysis of the factors influencing the effectiveness of the teacher's professional activity and the success of schoolchildren's teaching is carried out on the basis of the correlation coefficients of Pearson, Spearman and Kendall. To test the hypothesis about the equality of the general means of two independent samples, the Student's t-test was applied to the data. Revealed and investigated the laws governing the processes in the education system.

Key words: Bug Data, Dask, Pandas, educational process, correlation analysis, performance of the teacher

Введение

Повсеместное распространение технологий и доступа к Интернету привели к экспоненциальному росту объемов генерируемых данных. Благодаря огромным массивам данных у организаций появляются дополнительные возможности для роста и повышения качества оказываемых услуг. (см. [1]) Однако остро встает вопрос о том, как управлять, анализировать и извлекать ценность из "сырых" данных. Ежегодно отмечается рост числа организаций, которые начинают использовать алгоритмы математической обработки, технологий BigData и машинного обучения для повышения эффективности принимаемых решений, персонализации опыта клиентов, оптимизации процессов и генерации стратегических идей на основе анализа больших данных. (см. [1, 3]) Технологии BigData позволяют обработать большой объем неструктурированных данных, систематизировать их, проанализировать и выявить закономерности там, где человеческий мозг никогда бы их не заметил. Это открывает совершенно новые возможности по использованию данных. (см. [2]) Сегодня без Big Data управлять мегаполисом практически невозможно. Речь идёт о транспортных потоках, здравоохранении, внедрении системы искусственного интеллекта, градостроительных решениях, предотвращении правонарушений и в ещё целом ряде аспектов. Образование с этой точки зрения является наиболее сложной сферой. (см. [4]) К настоящему времени в мире анализ больших данных и применение технологий машинного обучения рассматривается в отрасли образования как критическая технология, которая позволит произвести важнейший переход к доказательной образовательной политике и к доказательной педагогике, а также осуществить индивидуализацию образовательного процесса через учет личностных особенностей, учебных и социально-экономических факторов, влияющих на эффективность профессиональной деятельности учителя и успешность школьника. (см. [5])

В этой работе продемонстрировано применение методов BigData для проведения анализа факторов влияющих на успешность школьного учителя. Анализ проведен на основе набора данных собранных в системе "Электронное образование для Республики Татарстан" за период с 2015 по 2020 годы.

1. Библиотеки Pandas и Dask для исследовательского анализа больших данных В настоящее время сбора и анализа больших данных всё чаще используется язык программирования Python. (см. [7]) Связано это прежде всего, с простотой языка, а также доступностью и разнообразием современных библиотек. Он предлагает богатый набор библиотек для общего анализа данных, для решения таких задач, как научные вычисления, визуализация данных и т. п. Один из лучших инструментов для разведочного анализа данных в Python - это библиотека Pandas. Эта высокоуровневая библиотека позволяет строить сводные таблицы, выделять колонки, использовать фильтры, выполнять группировку по параметрам, запускать функции (сложение, нахождение медианы, среднего, минимального, максимального значений), объединять таблицы и многое другое. (см. [6])

Однако Pandas не поддерживает механизмы параллельной обработки данных, поэтому анализ данных может занимать слишком долгое время (см. [10]). В результате этому пакету не удаётся на полную мощность воспользоваться возможностями современных многоядерных процессоров. Более

совершенный инструмент для обработки больших данных это пакет Dask. Dask - гибкая библиотека параллельных вычислений для аналитики, предназначенная главным образом для обеспечения масштабируемости и расширения возможностей существующих пакетов и библиотек. (см. [8]) Это самый революционный инструмент для обработки данных. Он поддерживает структуры данных Pandas DataFrame и Numpy Array и может быть запущен на локальном компьютере или масштабироваться для запуска в кластере. Dask позволяет с минимальными изменениями кода запустить программу параллельно, используя преимущества вычислительной мощности. Можно даже сконфигурировать Dask для использования ресурсов тысячи машин - каждой с несколькими ядрами. Одной из самых важных особенностей Dask является его удобство в использовании. Dask предоставляет панель мониторинга в режиме реального времени, которая выделяет ключевые метрики задачи обработки, выполняемой пользователем, такие как текущий прогресс проекта, потребление памяти и многое другое (см. рис 1).

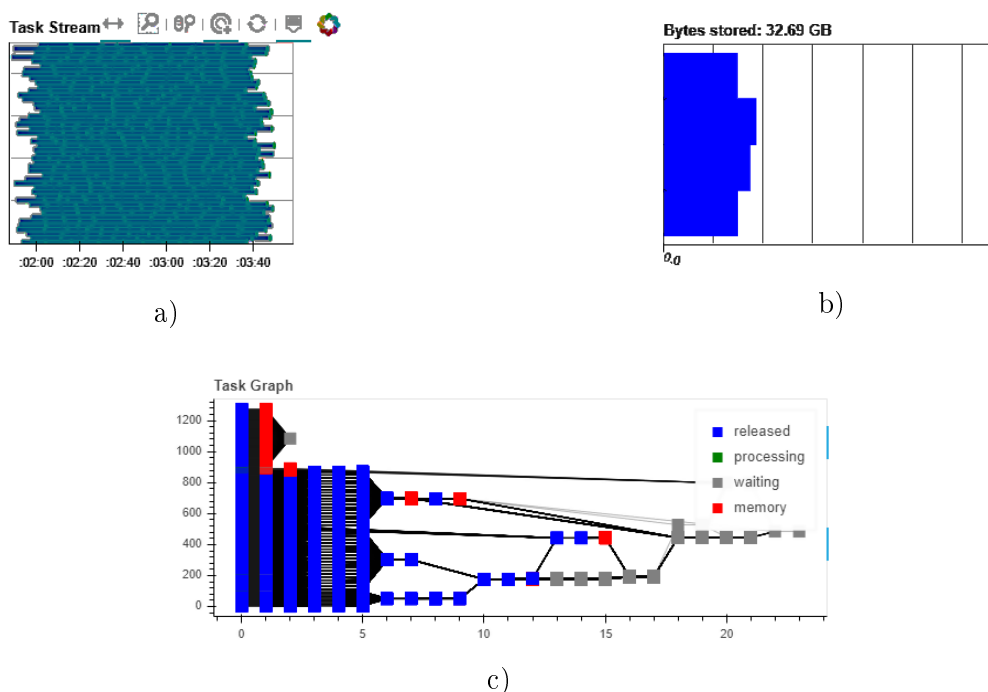


Рис. 1: Панель мониторинга: а) Dask Task Stream; б) Dask Memory; в) Task Graph Dask.

2. Построение информационно-аналитической системы для проведения интеллектуального анализа данных об образовательном процессе, на основе массивов больших данных.

С помощью описанных выше библиотек Dask и Pandas нами была построена информационно-аналитическая система для обработки и изучения массивов больших данных о деятельности педагога, администрации школы и успешности учащихся. Основу исследования составили данные, собранные через Государственную информационную систему "Электронное образование в Республике Татарстан". Система включает в себя базы данных образовательной информации по всем учащимся и всем педагогам общеобразовательных организаций РТ. Для исследования наборы данных предоставлялись в формате Comma-separated values (CSV файлов) общим объемом около 60 ГБ.

В обработку поступила следующая информация (с 2015 по 2020 года):

- По ученикам
 - Текущая успеваемость учащихся средней школы (с 1 по 11 классы) по Республике Татарстан .
- По учителям
 - Данные по социально-демографическим характеристикам, данные о прохождении педагогической аттестации, данные об образовании, данные, характеризующие профессиональную деятельность педагогов, включая темы уроков, задаваемые домашние задания, данные по учебной нагрузке, расписанию, преподаваемым предметам всех педагогов общеобразовательных организаций Республики Татарстан.

Для загрузки данных из CSV-файлов в Dataframe Dask-кластера использовалась функция `dask.dataframe.read_csv()`. На рисунке 2 изображен фрагмент загруженных данных.

| | edu_subject_id | title | short_title |
|---|----------------|---------------------------|-------------------|
| 0 | 1 | Русский язык | Рус.яз. |
| 1 | 2 | Литература | Лит-ра |
| 2 | 3 | Литературное чтение | Лит. чтение |
| 3 | 4 | Татарский язык | Тат. язык |
| 4 | 5 | Татарская литература | Тат. лит-ра |
| 5 | 6 | Литературное чтение (тат) | Лит. чтение (тат) |
| 6 | 7 | Английский | Англ. яз |
| 7 | 8 | Немецкий | Нем. яз. |
| 8 | 9 | Французский | Франц. язык |
| 9 | 10 | Чувашский | Чувашский |

Рис. 2: Вывод фрагмента результата DataFrame "edusubjects.csv"

Данные описывающие разные сущности (оценки, темы уроков, расписание, информация о педагогах и учениках) находятся в отдельных *csv*- файлах. Для проведения анализ часто приходится применять операции объединения (слияния) сущностей по какому-либо ключу. В Dask для объединения фреймов данных применялся метод `DataFrame.merge()`.

```
subjects=school_subjects.merge(edu_subjects, on='edu_subject_id', how='inner')
subjects.head()
```

Рис. 3: Объединение фреймов данных *school_subjects* и *edu_subjects* по столбцу *edu_subject_id*

Тип выполняемого слияния (параметр `how`) может принимать один из следующих значений:

- `left`: аналогично SQL `left outer join`;
- `right`: аналогично SQL `right outer join`;
- `outer`: аналогично SQL `full outer join`;
- `inner`: аналогично SQL `inner join`.

| | subject_id | edu_subject_id | title | short_title |
|---|------------|----------------|------------|-------------|
| 0 | 2 | 2 | Литература | Лит-ра |
| 1 | 30 | 2 | Литература | Лит-ра |

Рис. 4: Фрагмент результата слияния фреймов данных *school_subjects* и *edu_subjects* по столбцу *edu_subject_id*

С помощью описанных выше методов все полученные данные о деятельности педагога, администрации школы и успешности учащихся были обработаны, структурированы и объединены в общий DataFrame, к которому применялись методы математической статистики. В качестве критерия успешности учителя был взят уровень обученности его учеников, который рассчитывался как средняя годовая оценка в каждом классе по каждому предмету.

3. Корреляционный анализ с расчётом коэффициентов корреляции Пирсона, Спирмена и Кендалла.

Для изучения корреляции между переменными был применен корреляционный анализ. На основе этого метода сравнивались коэффициенты корреляции между одной парой или множеством пар признаков для установления между ними статистических взаимосвязей. К настоящему времени разработано великое множество различных коэффициентов корреляции. Однако самые важные меры связи - Пирсона, Спирмена и Кендалла. (см. [9]) С использованием этих коэффициентов корреляции и был произведен корреляционный анализ факторов, влияющих на эффективность профессиональной деятельности учителя и успешность обучения школьника.

3.1 Коэффициент корреляции Пирсона

Критерий корреляции Пирсона - это метод параметрической статистики, позволяющий определить наличие или отсутствие линейной связи между двумя количественными показателями, а также оценить ее тесноту и статистическую значимость. Для определения взаимосвязи двух переменных корреляционный анализ с расчетом коэффициентов корреляции Пирсона, вычисляется по следующей формуле:

$$c_{xy} = \frac{\sum(x_i - x)(y_i - y)}{\sqrt{\sum(x_i - x)^2 \sum(y_i - y)^2}}$$

, где x_i - значения, принимаемые в выборке X, y_i - значения, принимаемые в выборке Y;

3.2 Коэффициент корреляции Спирмена

Коэффициент ранговой корреляции Спирмена - это количественная оценка статистического изучения связи между явлениями, используемая в непараметрических методах. Коэффициенты корреляции Спирмена вычисляются по формуле:

$$p = 1 - \frac{6\sum d^2}{n^3 - n}$$

, где d^2 - квадратов разностей между рангами; N - количество признаков, участвовавших в ранжировании.

3.3 Коэффициент корреляции Кендалла

Коэффициент ранговой корреляции τ -Кендалла является альтернативой методу определения корреляции τ -Спирмана. Он предназначен для определения взаимосвязи между двумя ранговыми переменными. Интерпретация результатов вычисления коэффициент ранговой корреляции τ -Кендалла определяется как разность вероятностей совпадения и инверсии в рангах. Коэффициенты

корреляции Кендалла рассчитываются по следующей формуле:

$$\tau = \frac{P(p) - P(q)}{\frac{N(N-1)}{2}},$$

где $P(p)$ - число совпадений, $P(q)$ - число инверсий, N - объем выборки.

Анализ значений коэффициентов корреляции показал что коэффициентов корреляции Спирмена дает наивысший значения. При вычислении взаимосвязи двух переменных квалификационной категории и средней оценки учителя в старших классах (5-11класс) была выявлена слабая положительная корреляционная связь по русскому языку, литературе, английскому языку и средняя положительная корреляция по татарскому языку и татарской литературе. Средние баллы учеников растут с повышением категории учителей (Таблица 1). Все значения коэффициента корреляции статистически достоверны ($p\text{-value} < 0.05$). Таким образом, система присвоения категорий в целом справедлива, поскольку обучающиеся у учителей с более высокой категорией достигли в среднем более высоких результатов.

Таблица 1: Корреляционный анализ с расчетом коэффициентов корреляции г-Спирмена для определения взаимосвязи двух переменных квалификационной категории и средней оценки

| Год\Предмет | Рус. яз | Лит-ра | Англ. яз. | Тат. яз. | Тат. лит-ра |
|-------------|---------|--------|-----------|----------|-------------|
| 2015-2016 | 0.15 | 0.15 | 0.11 | 0.093 | 0.14 |
| 2016-2017 | 0.14 | 0.14 | 0.16 | 0.15 | 0.13 |
| 2017-2018 | 0.14 | 0.13 | 0.14 | 0.17 | 0.17 |
| 2018-2019 | 0.12 | 0.12 | 0.12 | 0.33 | 0.27 |
| 2019-2020 | 0.12 | 0.12 | 0.16 | 0.33 | 0.24 |

При нахождении корреляции между выборками X и Y , где X - общий стаж работы учителя, Y - средняя оценка учителя по предмету была обнаружена слабая отрицательная корреляция по таким предметам как ОБЖ, Физкультуре, Физике и ИЗО (Таблица № 2).

Таблица 2: Корреляционный анализ с расчетом коэффициентов корреляции г-Спирмена для определения взаимосвязи двух переменных общего стажа работы учителя и средней оценки

| Год\Предмет | ОБЖ | P-value | Физ-ра | P-value | Физика | P-value | ИЗО | P-value |
|-------------|--------|---------|--------|---------|--------|---------|--------|---------|
| 2015-2016 | -0.11 | 0 | -0.082 | 0 | -0.091 | 0.0002 | -0.11 | 0.0024 |
| 2016-2017 | -0.12 | 0 | -0.12 | 0 | -0.07 | 0.0061 | -0.091 | 0.0084 |
| 2017-2018 | -0.12 | 0 | -0.11 | 0 | -0.1 | 0.0001 | -0.14 | 0 |
| 2018-2019 | -0.091 | 0.014 | -0.093 | 0 | -0.11 | 0 | -0.14 | 0 |
| 2019-2020 | -0.13 | 0 | -0.11 | 0 | -0.075 | 0.0041 | -0.14 | 0 |

При рассмотрении взаимосвязи возраста педагога и средней оценки была обнаружена слабая отрицательная корреляционная зависимость. У педагогов младшего возраста по отношению к их коллегам старшего возраста более высокие показатели средней оценки (Таблица № 3)

Таблица 3: Корреляционный анализ с расчетом коэффициентов корреляции Пирсона для определения взаимосвязи двух переменных возраста педагога и средней оценки ((p-value<0.05))

| Год\Предмет | Физ-ра | Физика | Информ. и ИКТ | ОБЖ | ИЗО |
|-------------|--------|--------|---------------|--------|--------|
| 2015-2016 | -0.074 | -0.083 | -0.097 | -0.12 | -0.095 |
| 2016-2017 | -0.11 | -0.07 | -0.09 | -0.1 | -0.1 |
| 2017-2018 | -0.11 | -0.088 | -0.074 | -0.092 | -0.15 |
| 2018-2019 | -0.096 | -0.1 | -0.097 | -0.077 | -0.13 |
| 2019-2020 | -0.11 | -0.08 | -0.12 | -0.14 | -0.14 |

3.4 Расчет t-критерия Стьюдента при сравнении средних величин

Для проверки гипотезы о равенстве генеральных средних двух независимых выборок к данным был применен t-критерий Стьюдента. Для сравнения средних величин t-критерий Стьюдента рассчитывался по следующей формуле:

$$t = \frac{M_1 - M_2}{\sqrt{m_1^2 + m_2^2}},$$

где M_1 - средняя арифметическая первой сравниваемой совокупности (группы), M_2 - средняя арифметическая второй сравниваемой совокупности (группы), m_1 - средняя ошибка первой средней арифметической, m_2 - средняя ошибка второй средней арифметической.

Исследование t-критерия Стьюдента для учителей - мужчин и учителей - женщин показало, что учителя - мужчины ставят оценку выше, нежели учителя - женщины. Из этого можно сделать вывод, что женщины педагоги более требовательны к ученикам и критично оценивают отношение учеников у учебного процесса.

Таблица 4: t-критерий Стьюдента для сравнения учителей-мужчин и учителей-женщин

| Год\Предмет | Математика | | P-value | География | | P-value |
|-------------|------------|------|----------|-----------|------|----------|
| | муж. | жен. | | муж. | жен. | |
| 2015-2016 | 3.77 | 3.71 | 1,61E-05 | 4.06 | 3.98 | 7,02E-08 |
| 2016-2017 | 3.79 | 3.72 | 3,65E-06 | 4.07 | 3.97 | 1,66E-11 |
| 2017-2018 | 3.82 | 3.74 | 1,57E-07 | 4.06 | 3.95 | 2,62E-10 |
| 2018-2019 | 3.86 | 3.76 | 2,78E-09 | 4.06 | 3.96 | 2,84E-10 |
| 2019-2020 | 3.9 | 3.81 | 1,59E-06 | 4.06 | 3.96 | 3,44E-10 |

Сравнение учителей по общему стажу работы показало, что учителя со стажем работы менее 25 лет ставят по всем предметам отметки выше, чем учителя со стажем 25 лет и более. Исключением является английский язык.

Таблица 5: t-критерий Стьюдента для сравнения учителей со стажем работы менее 25 лет и учителей со стажем более 25 лет

| Год\Предмет | История | | P-value | Обществознание | | P-value |
|-------------|-----------|-----------|---------|----------------|-----------|---------|
| | Стаж > 25 | Стаж < 25 | | Стаж > 25 | Стаж < 25 | |
| 2015-2016 | 3.99 | 3.96 | 0,004 | 4.05 | 4.01 | 0,00234 |
| 2016-2017 | 3.98 | 3.94 | 0,0004 | 4.05 | 4.01 | 2,8E-05 |
| 2017-2018 | 3.98 | 3.96 | 0,0188 | 4.05 | 4.01 | 0,00052 |
| 2018-2019 | 3.99 | 3.95 | 0,0001 | 4.04 | 4.0 | 0,00011 |
| 2019-2020 | 4.04 | 3.97 | 3,5E-07 | 4.03 | 3.98 | 3,2E-05 |

Таблица 6: t-критерий Стьюдента для сравнения учителей со стажем работы менее 25 лет и учителей со стажем более 25 лет по английскому языку

| Год\Предмет | Английский язык | | P-value |
|-------------|-----------------|-----------|----------|
| | Стаж > 25 | Стаж < 25 | |
| 2015-2016 | 3.87 | 3.9 | 0,017139 |
| 2016-2017 | 3.87 | 3.89 | 0,032576 |
| 2017-2018 | 3.88 | 3.92 | 0,000131 |
| 2018-2019 | 3.89 | 3.92 | 0,008193 |
| 2019-2020 | 3.9 | 3.93 | 0,022467 |

По возрастным характеристикам исследование показало, что у учителей моложе 50 лет отметки по всем предметам выше, чем у учителей чей возраст превышает 50 лет. Исключением является английский язык. По английскому языку у молодых учителей уровень оценок учеников выше, нежели у педагогов старшего возраста.

Таблица 7: t-критерий Стьюдента для сравнения учителей моложе 50 лет и учителей старше 50 лет

| Год\Предмет | Математика | | P-value | Физкультура | | P-value |
|-------------|------------|-----------|----------|-------------|-----------|----------|
| | Возр < 50 | Возр > 50 | | Возр < 50 | Возр > 50 | |
| 2015-2016 | 3.73 | 3.71 | 0,007463 | 4.52 | 4.48 | 0,001506 |
| 2016-2017 | 3.74 | 3.72 | 0,016869 | 4.52 | 4.46 | 7,88E-06 |
| 2017-2018 | 3.77 | 3.73 | 4,68E-05 | 4.52 | 4.47 | 3,01E-06 |
| 2018-2019 | 3.79 | 3.76 | 0,000402 | 4.52 | 4.47 | 0,00014 |
| 2019-2020 | 3.84 | 3.8 | 0,000514 | 4.54 | 4.49 | 0,000242 |

При сравнении учителей по квалификационной категории выявилось, что у педагогов без квалификационной категории по русскому языку, английскому языку и литературе оценки по предметам ниже, чем у педагогов с высшей квалификационной категорией.

Таблица 8: t-критерий Стьюдента для сравнения без квалификационной категории оценки и педагогов с высшей квалификационной категорией

| Год\Предмет | Рус. яз. | | P-value | Лит-ра | | P-value |
|-------------|------------|----------|----------|------------|----------|----------|
| | Высш. кат. | Нет кат. | | Высш. кат. | Нет кат. | |
| 2015-2016 | 3.68 | 3.81 | 2,0E-09 | 3.95 | 4.09 | 3,08E-08 |
| 2016-2017 | 3.68 | 3.81 | 6,1E-09 | 3.97 | 4.09 | 3,53E-06 |
| 2017-2018 | 3.73 | 3.82 | 4,49E-06 | 4.03 | 4.1 | 0,000147 |
| 2018-2019 | 3.74 | 3.81 | 3,5E-07 | 4.03 | 4.1 | 4,96E-06 |
| 2019-2020 | 3.74 | 3.8 | 1,5E-05 | 4.03 | 4.1 | 2,1E-05 |

Существенных различий по средним оценкам учителей с позицией "Учитель - классный руководитель" и учителей с позицией "Учитель - предметник" обнаружено не было.

3. Заключение. Разработана информационно-аналитическая система на основе библиотеки для проведения параллельных вычислений в кластерных вычислительных системах Dask. Проведен корреляционный анализ факторов влияющих на эффективность профессиональной деятельности учителя и успешность обучения школьника. Для проверки гипотезы о равенстве генеральных средних двух независимых выборок к данным был применен t-критерий Стьюдента. По результатам проведенного исследования, посвященного определению факторов, влияющих на эффективность профессиональной деятельности учителя и успешность обучения школьника, получены следующие выводы. Средний балл учителя зависит от пола, возраста, квалификационной категории и стажа работы, и не зависит от профессиональной позиции ("Учитель - классный руководитель" и "Учитель - предметник").

Л И Т Е Р А Т У Р А

1. Агеева Е. В. , Афанасова М. А., Баландина А. С., Балашова Н. В., Баннова К. А. Цифровизация финансово-кредитной сферы в современной России, Директ-Медиа, 2019.
2. Веретенников А. В. BigData: анализ больших данных сегодня / А. В. Веретенников. - Текст : непосредственный // Молодой ученый. - 2017. - № 32 (166). - С. 9-12. - URL: <https://moluch.ru/archive/166/45354/> (Дата обращения: 05.09.2020)
3. Кельчевская Н.Р., Ширинкина Е.В. Региональные детерминанты эффективного использования человеческого капитала в цифровой экономике // Экономика региона. - 2019. - Т.15, вып.2, - С.465 - 482
4. Лобанов Н. Б. Логистика и мегаполис: проблемы и решения [Электронный ресурс] URL: <http://www.kpilib.ru/article.php?page=636> (Дата обращения: 17.09.2020)
5. Buniyamin N., Mat U. B., Arshad P. M. Educational data mining for prediction and classification of engineering students achievement // Paper presented at the 2015 IEEE 7th International Conference on Engineering Education, ICEED 2015. 49-53.
6. Хаштамов А. Статья "Введение в Pandas: анализ данных на Python"[Электронный ресурс] // URL: <https://khashtamov.com/ru/pandas-introduction/> (Дата обращения: 12.09.2020)
7. Автор Статья "Анализ данных с использованием Python"[электронный ресурс] // URL: <https://habr.com/ru/post/353050/> (Дата обращения: 09.09.2020)
8. Varangaonkar A. Rocklin M. Dask: Parallel Computation with Blocked algorithms and Task Scheduling, Proceedings of the 14th Python in Science Conference, 130 - 136 ,2015
9. Харченко М.А. Корреляционный анализ: Учебное пособие для вузов. - Воронеж: Изд-во ВГУ, 2008. - 31 с.

10. **Абдрахманов М. И.** Pandas. Работа с данными. 2-е изд. - Издательство: Devpractice Team, 2020. - 171 с.

R E F E R E N C E S

1. **Ageeva E. V. , Afanasova M. A., Baladina A. S., Balashova N. V., Banannova K. A.** Digitalization of the financial and credit sphere in modern Russia, Direct Media, 2019.(in Russian)
2. **A. V. Veretennikov** BigData: big data analysis today / A. V. Veretennikov. - Text: direct // Young scientist. - 2017. - № 32 (166). - p. 9-12. - URL: <https://moluch.ru/archive/166/45354/> (in Russian)
3. **Kelchevskaya, N.R., Shirinkina, A.V.** Regional determinants of effective use of human capital in the digital economy // Economy of the region. - 2019. - Т.15, вып.2, - p.465 - 482 (in Russian)
4. **Lobanov N.B.** Logistics and megapolis: problems and solutions [Electronic resource] URL:<http://www.kpilib.ru/article.php?page=636> (in Russian)
5. **Buniyamin N., Mat U. B., Arshad P. M.** Educational data mining for prediction and classification of engineering students achievement // Paper presented at the 2015 IEEE 7th International Conference on Engineering Education, ICEED 2015. 49-53.
6. **Khashtamov A.** Introduction to Pandas: Data Analysis in Python [Electronic resource] // URL: <https://khashtamov.com/ru/pandas-introduction/> (Date of the application: 12.09.2020)
7. **Автор** Data Analysis Using Python [electronic resource] // URL: <https://habr.com/ru/post/353050/> (Date of the application: 09.09.2020)
8. **Varangaonkar A. Rocklin M.** Dask: Parallel Computation with Blocked algorithms and Task Scheduling, Proceedings of the 14th Python in Science Conference, 130 - 136 ,2015
9. **Kharchenko M.A.** Correlation analysis analysis: A useful tool for learning. - Voronez: VSU, 2008. - 31 с. (in Russian)
10. **Abdrakhmanov M.I.** Pandas. Work with data. 2nd edition. - Publisher: Devpractice Team, 2020. - p. 171

Содержание

Музафарова А.И., Гафаров Ф.М. *Технологии больших данных для анализа влияния компонентов образовательной среды на результативность профессиональной деятельности учителей* 1