Monitoring and Calibrating the Efficiency of the Remote and Hybrid Remote Work

OKSANA NIKIFOROVA[1], VITALY ZABINIAKO[2], PĀVELS GARKALNS[1],
JURIJS KORNIENKO[3], VLADIMIRS NIKULSINS[2], ANDREJS ROMANOVS[1]

[1]Faculty of Computer Science and Information Technology,
Riga Technical University,
10 Zunda kanals, Riga, LV-1048,
LATVIA

[2]Research and Development Department,
"ABC software" Ltd.,
Tallinas iela 51a, Riga, LV-1012,
LATVIA

[3]Microsoft Solutions Department,
"ABC software" Ltd.,
Tallinas iela 51a, Riga, LV-1012,
LATVIA

Abstract: - Remote (or hybrid remote) work is currently becoming more and more trending. This happened in recent years, driven by technological advancements and the global COVID-19 pandemic. This model of work offers immense opportunities for employers and employees, such as increased flexibility, improved work-life balance, and reduced commute time. However, the shift to remote work also comes with its own set of challenges and complexities. The paper describes the results of the research, which specifies data for the approbation of the method for calibrating the efficiency of remotely working employees and observing the dynamics of work productivity in an experimental environment developed during this research. The method offered for efficiency calibration supposes to use data regarding workers' activities according to business domain.

Key-Words: - Remote work, work efficiency, software development effort estimation, story points, velocity, behavioral pattern, intelligent control, data mining

Received: August 9, 2022. Revised: March 23, 2023. Accepted: April 15, 2023. Published: May 22, 2023.

1 Introduction

Remote work is the area that has grown rapidly in the last couple of years, and it comes with its own timekeeping issues, [1], [2], [3], [4]. Theoretically, there should not be any significant differences, in working in the office or remotely, but it can be observed that such differences do exist in practice because, in the case of remote work, employers want not only to precisely account for the time worked, but also – to be sure of the efficiency of the tasks being performed by employees.

It is necessary to calibrate such efficiency, especially in the modern situation, when work is organized in a hybrid way – by working both from home and from an office space, as needed.

The environment for determining the efficiency of employees is the usage of one or more information systems (IS-s) of different types, within which the employee's job duties are being performed. This process can take place both from home (isolated from the work environment and the collective of colleagues), and also – in office premises. Another aspect is related to the fact that an employee can work both with only one IS and with several IS-s in parallel, which requires the simultaneous integration of several data sources and the parallel reading and analysis of relevant audit data.

Also, it is expected that the metrics of employees' efficiency parameters and calculations are compared with information from project management software tools, in which the employee reports on the progress of completion of his tasks. Consequently, the development of this method requires the acquisition and design of a specifically organized set of data, so that it can be used in the future in the method for calibrating the efficiency of employees, which was developed during this study.

The goal of the research is to develop a data mining-based approach that will evaluate the efficiency of the company's remote employees who use IS-s for work, based on the efficiency profiles of the users themselves and the corresponding user groups.

In the course of the research, the development of new methods and the relevant theoretical base was carried out, allowing to use of these together with a data mining-based approach, assessing work efficiency and its dynamics of changes for on-site/remote workers (in different time periods).

Appropriate methods provide the employer with the opportunity to monitor and respond to potential unwanted deviations from the expected and planned work volumes and efficiency. At the same time, it allows the employer to offer the employee possible corrective actions (if necessary), for example – additional motivation and training, protection from burnout syndrome, as well as – providing instructions regarding insufficient self-discipline during remote work, if such cases do arise. Also, the employee will be able to review the scope and productivity of his work on a daily/weekly/monthly/yearly basis.

As a result, within the framework of the research, a solution was developed that allows the business management of various industries to monitor the efficiency of employees, as well as to improve it by making reasonable operational decisions, with the help of data mining-based analytical methods. The solution provides an opportunity to both identify and correct the "ups and downs" of the efficiency of a given employee, as well as to identify and influence relevant cases for multiple employees within similar qualifications/specialization/role/job positions.

The paper is organized as follows: the second chapter outlines the background and current problems of remote work of IS users, the third chapter defines the proposed method for solving mentioned non-trivial task, and the last chapter provides conclusions and outlines the future research capabilities.

2 Background of the Area of Remote Work Monitoring and Automatic Control

Taking into consideration the current world health situation and the relevant work organization trends, which have been affected by the COVID-19 pandemic (and may be affected by other pandemics in the future), it is clear that the emphasis on self-isolation and transferring work activities to home has a significant impact on the intensity of the use of IT systems. The same concerns also habit the performance of employees' direct job duties and also have a great impact on other aspects of users and IS-s usage.

The rapidly increasing number of remotely working users, on the one hand, reduces the risks of spreading the epidemic and improves the overall environmental ecology (significantly reducing CO_2 emissions and traffic jams in cities) and the maintenance costs of the working environments (office premises, etc.). But on the other hand, it affects the efficiency of the daily workflow, as well as complicates management and operational monitoring of the activities that users perform in the digital environment during the performance of their direct work duties, so this must be managed accordingly, [5]. Factors that can negatively affect the efficiency of remote workers are:

- Decreasing work motivation when working remotely and alone for a long time period.
- Increased risk of employee "burnout", because the feeling that one must work alone in the performance of work duties prevails, there is a lack of direct support from colleagues, there is an "intrusion" of work matters into the space of private life, which increases the level of stress and contributes to this "burnout" in the long term.
- Negative influence from various factors of private life, such as the performance of various household chores, distractions from work by family members, etc.
- Limited opportunities for acquiring new knowledge and growth, especially for new employees with insufficient experience, compared to joint work and growth opportunities in the office, while working in a unified team.

Some additional important aspects and problems of remote work are as follows following:

- Remote work requires an enhanced focus on collaboration and communication practices. For remote teams, establishing daily check-ins, and regular meetings, and creating virtual collaboration channels becomes crucial for facilitating efficient communication. Without such practices in place, remote workers may feel isolated and disconnected, which can negatively impact work productivity.

- There is a greater need for companies to invest in the right technological infrastructure to facilitate remote work. This includes providing the required hardware, software, and information security protocols. Ensuring that employees have access to secure and reliable network connections is also crucial for efficient and productive remote work.
- There is a need for employers to implement clear policies and procedures to guide remote work practices. This includes setting clear expectations, establishing remote work protocols, and developing clear communication channels for remote teams. Clarity is vital when it comes to remote work, as it leaves no room for confusion, misunderstandings, or misinterpretation.
- Remote work creates challenges for organizations in terms of employee engagement and efficiency. Without the physical office space and interaction of face-to-face meetings, remote employees may feel less engaged and less committed to their work. Managers must be proactive in fostering employee engagement through regular check-ins, virtual team-building activities, and employee feedback programs.

Consequently, there is a need to emphasize automated tools and means for employee efficiency evaluation, the usage and analysis of results of which do not depend on physical presence “on-site in the office”. Such an approach is significantly more efficient and even mandatory, especially in the world affected by COVID-19, where the activities of IS users can no longer be directly observed in person.

This inevitably forces us to turn to the involvement of machine learning and data mining methods in the workflow of monitoring daily user activities, to have the opportunity to evaluate employee efficiency, workability, and motivation drops.

To sum up – remote work has become a vital aspect of modern work culture. However, it requires significant effort, investment, and strategy to make it work efficiently. Companies need to create an environment that fosters collaboration, invests in technological infrastructure, creates a conducive work environment, and foster engagement and maintenance of efficient work of their employees. Addressing these challenges will help all organizations to thrive in the age of remote work.

This form of work produces several additional challenges, as there is currently no universal solution for assessing and optimizing employee efficiency in such an environment. The World has been working in this mode for less than three years, and in this regard – we propose to consider adopting a set of metrics from software development methods, [6], which have already been successfully applied in a hybrid work context.

We present our solution in this paper, which demonstrates how to calibrate employee efficiency for hybrid work environments. We used audit (log) data from one particular IT company and the work of according employees in their IS-s, which was collected over the time period of 10 months, to build our method for evaluating employee efficiency.

Our method claims universality, as we have already successfully applied its first, narrower version, [6], [7], [8], [9], for monitoring system engineers working with the AutoCAD design tool. The conclusion was, that the offered idea for efficiency metrics and the approach to calculate these and use data mining methods, is quite universal and can be applied to any business domain, where it is possible to collect data about activities performed by workers in the IS-s and tools for their duties.

Our approach is based on the application of well-based mathematical and analytical methods and metrics, which allow us to extract and evaluate useful information from the collected audit data. We offer our method as one of the solutions to this problem in hybrid work and believe that it could be successfully applied in different other organizations and companies.

3 Approach to Calibrate Work Efficiency based on Audit Data of Workers' Activities

In the scope of the research, an innovative method for calibrating the work efficiency of employees (who are, essentially users of various IS-s) is being developed.

3.1 The Conceptual Idea of the Approach

The method is based on the analysis of the behavior patterns of users of these IS-s. A set of new methods and a relevant theoretical base was developed, allowing to use of these together with a data mining-based approach, assessing work efficiency and its dynamics of changes for on-site/remote workers (in different time periods).

Appropriate methods provide the employer with the opportunity to efficiently monitor and respond to potential unwanted deviations from the expected and planned work volumes. At the same time, it allows the employer to offer the employee possible corrective actions (if needed).

By collecting the studied information about the work of users (employees) of several IS-s, a set of data that can characterize the work in the context of measuring the efficiency contains the following values:

- Actions (with a link to a specific multi-system user and the documents/information units he/she works with).
- Action start/end time, and thus – the duration of a given activity.
- The content of the work and its quality/contribution/business value performed in a given unit of time.
- In case it is possible to accumulate information about the distribution of employees by projects or departments – this information can also be used to calculate the efficiency metrics of the relevant projects or structural units.

The method for determining the metrics of employee efficiency parameters analyses the audit records of authenticated users of the IS from several message streams and looks for relationships between the actions performed in the user's IS and the change of states in a unit of time. By linking time with influencing factors and adding efficiency evaluation metrics using a data mining-based approach, an efficiency factor is determined that characterizes the analyzed user's behavior in the IS. Based on this determined efficiency factor, it is possible to evaluate the efficiency aspects of the work of the analyzed IS user (employee).

The approach offered by this study is capable of analyzing and determining user behavior in online mode. The analysis requires a set of data regarding users registered in the IS (such as their unique IDs), as well as data on users' activities from IS audit logs. When the appropriate plug-in receives data from various activity logs of the target IS – the user's sessions are structured in the solution according to predefined user behavior patterns.

After structuring is complete, the user's activities are subjected to the efficiency metrics calculations defined in the next subsection. The overall workflow of the efficiency evaluation architecture is shown in Figure 1. The audit data collected in the activity logs of IS users were identified as the primary input data in the course of the research, which allow for obtaining precise information regarding actions that users perform while fulfilling their work duties (both within one IS and also – in several IS-s in parallel).

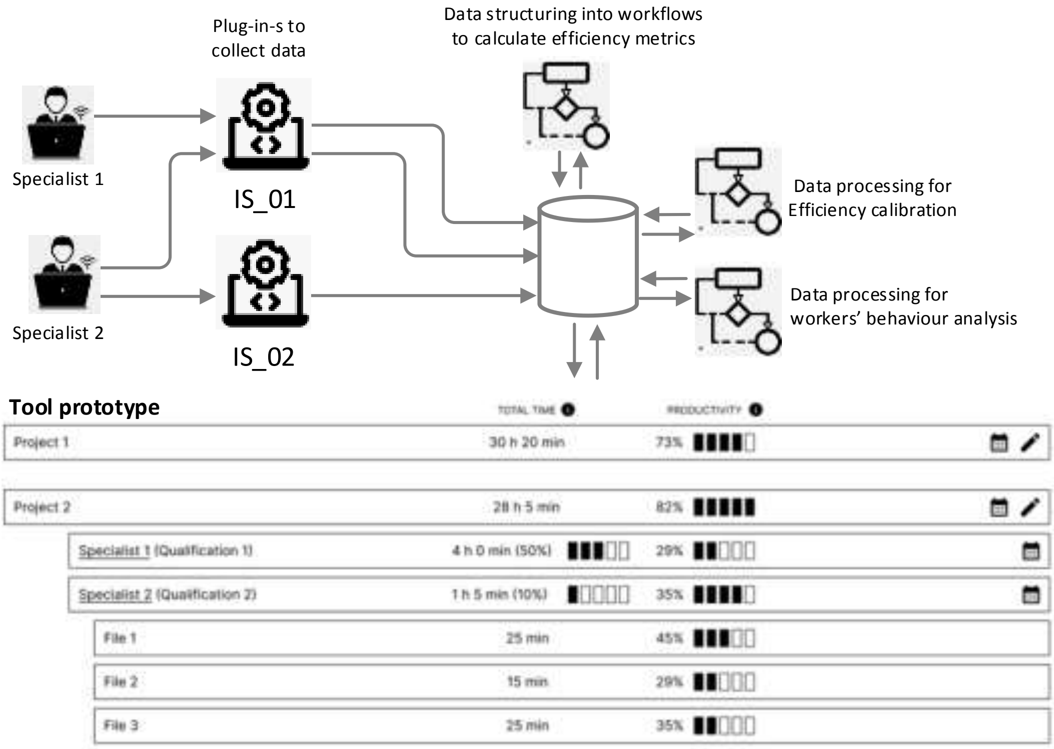


Fig. 1: The overall workflow of the data collection and tool prototype to support the approach for work efficiency calibration

The method of evaluating the efficiency of employees is based on the analysis of separate sets of metrics, where each set of metrics complements the other and allows obtaining a full-fledged evaluation of efficiency.

The first of such sets of metrics is related to the evaluation of efficiency in the context of the time (and accordingly – the work calendar). It consists of 3 metrics:

1. Time spent while working with an IS subsystem or file (the concept of workflow is defined, which corresponds to continuous work with the IS within a 5-minute time interval).
2. The velocity of executed actions, which is defined by experts in the related field and allows us to evaluate the contribution of the atomic executed action to the final goal of the work to be performed (this concept is borrowed from the methodology of software development, [6]).
3. A metric for evaluating the productivity of the end user, which corresponds to the sum of the velocity of executed actions divided by the length of the time interval of the respective workflow.

In general, this set of metrics allows us to get an idea of the IS user's contribution to the goal of the work to be performed in the context of time.

The second set of metrics is related to the behavioral patterns of IS users and their degree of efficiency. As part of the relevant analysis, action patterns that are characteristic to users are identified – i.e., characteristic sequences of actions, their duplicates, more / less frequently occurring chains of actions, unused actions, etc. This stage of analysis includes both simple efficiency evaluation phases (e.g. – the statistics of the most frequently executed actions and their execution times) and the more complex efficiency evaluation phases (e.g. – the data mining of the frequency of repeated chains).

In addition, the results of the mentioned sets of metrics are analyzed both for each user individually and also – for groups of users of IS-s, taking into consideration the logical assumption that activity patterns and average work efficiency of users with the same qualification levels should be correlated with each other.

The proposed method assumes that the dynamics of efficiency decline is considered significant only if it has been observed over a long period of time and the deviations from the benchmark are greater than should be allowed for the work to still be considered as efficient enough.

3.2 Preparation of the Initial Data Set

In general, in the scope of the research, the input data was defined, which allows for the evaluation of the efficiency of the company's working employees with the proposed metrics, which are used within the framework of the developed method.

The method bases the determination of work efficiency on the accumulated information on the usage of several IS-s in the scope of performance of work duties. The minimum amount of data that would be sufficient to determine the work efficiency of the authenticated user – is the names (identifiers) of the actions to be performed and execution time periods of these actions in different IS-s.

The evaluation of the general performance of the multi-system user (which is essentially the speed of execution of an action) requires information containing the following:

- The ID of the authenticated system user;
- A list of actions performed by this user in the files (databases/subsystems) defined in the tasks of a specific project (or department).

It is expected that in-office (or remotely) working employees daily work with several IS-s, and for the research results to be universal and applicable in any IT-related industry, the input data should contain the following information:

1. Employee's identifier;
2. Name of the performed action;
3. Action performance time;
4. IS (or a separate file) in which the action is being performed;
5. Project (or department) in the scope of which the execution of particular action has been performed.

For the development of the method, a data set has been used, in which information on operations performed by 16,280 users in two IS-s, being employees in 1,257 departments (and branches of these), was collected within 10 months time period.

All data has been anonymized so that employees' personal information is obfuscated with a randomly selected female name, and departments have been named with animals' names. Since it was initially planned to perform efficiency calibration within the projects carried out in the company, the dataset for experiments assumed that the actions performed by each employee in the department are actions performed within the framework of one project.

The spreadsheet in Figure 2 shows the statistics of the full data set. The first column lists employees (users of IS-s). The second column (in green color) contains the total number of activities performed by the relevant employee in 10 months. The following columns with months' names show how many actions have been performed in the according month. The last two columns show min/max working hours of each employee in a day. It can be seen that the employees are divided into two types – those who work around the clock, and those who work 8 hours a day.

All employees in the table in Figure 2 are arranged in descending order of the number of completed actions. The “TOP” employees, who have a higher number of performed actions (compared to others) and who work an 8-hour working day, are marked with a yellow background.

These are employees (7 individuals in total) who have been selected to be included in the initial data for the approbation of our method.

		2020						2021				Days in total					
UserName		May	June	July	Aug	Sept	Oct	Jan	Febr	Mar	Apr	211					
						Number of actions								Days			
	Total (sor	263426	283958	288508	300135	286864	270217	265764	259342	275681	207057	Average for user	Average in May	Days in May	Month Count	MinTime	MaxTime
Amelia	39839	5992	5930	5681	4748	4255	2922	1906	3288	2203	2914	23.60	300	20	10	00:00:09	23:59:59
Gianna	23161	4288	2929	4443	2689	2014	2679	869	1157	1105	988	13.72	268	16	10	00:00:14	23:59:56
Olivia	52946	5156	6842	6389	7150	6382	5782	5131	953	5329	3832	31.37	258	20	10	08:08:22	18:32:20
Layla	20056	2504	2243	2414	2800	2691	4263	2481	660			11.88	228	11	8	00:00:05	23:59:59
Evelyn	25329	3185	3233	2819	1969	3189	1829	8	2679	3799	2619	15.01	228	14	10	00:00:16	23:59:57
Emma	41665	5151	4667	4114	4301	4767	4787	5244	4943	1584	2107	24.68	224	23	10	00:33:49	22:03:46
Ella	22142	3332	1431	1911	2271	2765	2602	3116	1156	2094	1464	13.12	222	15	10	00:00:01	23:59:57
Avery	21310	2761	1863	1915	2423	2888	2325	424	1933	2849	1929	12.62	212	13	10	00:00:01	23:59:18
Luna	24420	2307	2509	1619	2218	2668	2330	3561	2105	2179	2924	14.47	210	11	10	00:00:20	23:59:58
Everleigh	8163	1414	2426	260	120	789	2395	759				4.84	202	7	7	00:00:07	23:59:52
Charlotte	40698	3967	4851	2765	4964	4147	1488	4562	4904	5289	3761	24.11	198	20	10	01:11:17	23:43:23
Isabella	27428	1735	3309	2602	2700	2722	1160	3471	3805	4345	1579	16.25	193	9	10	00:00:00	23:59:59
Stella	16391	1523		910	2446	2185	2010	1870	2048	1797	1602	9.71	190	8	9	00:00:20	23:59:24
Penelope	20462	2940	2008	2255	2031	578	1307	2511	2732	3223	877	12.12	184	16	10	00:00:00	23:59:58
Kennedy	11478	1836	1495	464	334	1581	280	1936	1685	1067	800	6.80	184	10	10	00:00:12	23:59:51
Eleanor	22150	2718	3232	2294	3191	2608	1162	734	3072	1723	1416	13.12	181	15	10	00:00:14	23:59:59
Madison	18588	2695	2554	2891	2307	1278	2253	2744	231	882	753	11.01	180	15	10	00:00:00	23:59:33
Isla	17799	2310	2707	918	2373	2233	1325		2183	2422	1328	10.54	178	13	9	00:00:00	23:59:57
Elena	14610	2523	1947	2983	3444	3017	4	157	206	224	105	8.66	168	15	10	00:00:04	23:59:55
Ava	32698	3184	3684	3995	3753	1228	3500	3853	4793	2053	2655	19.37	159	20	10	07:39:24	17:31:02
Riley	17191	2504	2012	846	2050	2307	2264	996	1597	1659	956	10.18	157	16	10	00:00:19	23:59:47

Fig. 2: The statistics of the full data set collected about workers' activities

A fragment of the initial data set, which was submitted for the approbation of the algorithm and metrics, is shown in Figure 3. The full data set contains 20,227 action records performed by seven employees over one month.

Employees work in four different departments. In addition to user and department data, the names of the executed actions have also been obfuscated (leaving the original names only for general actions, such as LOGIN, EXIT, MAIN, etc., but renaming other actions to “Action_001”, “Action_002”, etc.).

UserID	ClickTime	CommandName	Unit	Project
Sophia	2020-05-04T07:45:23.246+02:00	LOGIN	IISR	Donkeys
Sophia	2020-05-04T07:45:23.743+02:00	INICIALIZE TASK LIST	IISR	Donkeys
Sophia	2020-05-04T07:45:24.626+02:00	MAIN	IISR	Donkeys
Sophia	2020-05-04T07:45:28.000+02:00	Action_004	IISR	Donkeys
Sophia	2020-05-04T07:45:28.083+02:00	Action_004	IISR	Donkeys
Sophia	2020-05-04T07:45:47.000+02:00	Action_004	IISR	Donkeys
Sophia	2020-05-04T07:53:05.000+02:00	Action_004	IISR	Donkeys
Ava	2020-05-04T08:01:46.822+02:00	LOGIN	IISR	Flies
Ava	2020-05-04T08:01:47.468+02:00	INICIALIZE TASK LIST	IISR	Flies
Ava	2020-05-04T08:01:48.459+02:00	MAIN	IISR	Flies
Ava	2020-05-04T08:02:48.612+02:00	Action_005	IISR	Flies
Ava	2020-05-04T08:02:57.000+02:00	Action_005	IISR	Flies
Sophia	2020-05-04T08:03:27.000+02:00	Action_004	IISR	Donkeys
Ava	2020-05-04T08:03:43.939+02:00	Action_005	IISR	Flies
Ava	2020-05-04T08:03:44.568+02:00	MAIN	IISR	Flies

Fig. 3: Initial data fragment example

3.3 Definition of the Workflow for the Method Evaluation

In the proposed method, it was experimentally proven that if the break between the actions to be performed in 5 minutes or less, it can still be considered that continuous work is ongoing because the actions of an employee in performing tasks are not only related to “clicking buttons” in the IS but also – with thinking and management (administrative) activities. However, if the break between the execution of actions within one information system is 15 minutes long or more, then it can be considered that continuous work was no longer taking place.

An analysis of breaks between actions has been performed for all employees in the gathered experimental dataset, which confirmed that these 5 / 15-minute timeout intervals are valid assumptions for further data analysis.

Figure 4 shows an example of time breaks between actions for the employee Sophia.

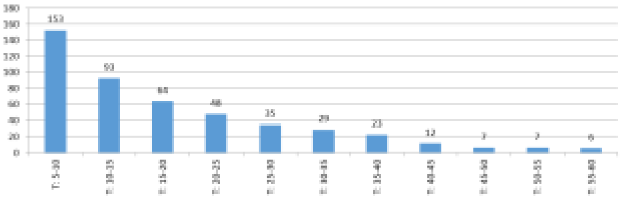


Fig. 4: Popularity of breaks for the employee Sophia

Taking into consideration the results of the experiments, the concept of "workflow" has been introduced, the essence of which is shown schematically in Figure 5.

By executing the first action in a specific IS, the start time of a certain workflow is recorded for this specific IS – it is shown with a black bar in Figure 5, and if a work interruption is registered with a pause of 5 minutes (or more) – it interrupts the workflow of actions performed in this IS. However, as a rule, if a gap of more than 5 minutes (but less than 15 minutes) is recorded between the execution of actions – the worker's overall workflow (shown by the grey bar in Figure 5), is still considered to be continuous.

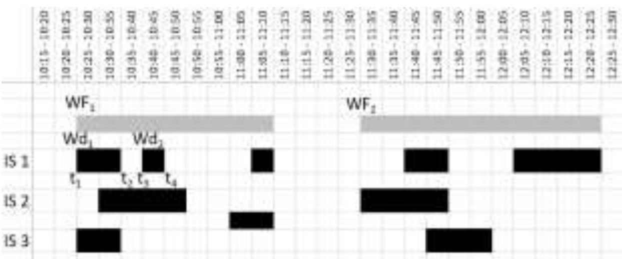


Fig. 5: Workflows and interruptions examples

Wd_i marks individual "small" jobs, while WF_i marks the overall workflow interval, which is a combination of individual "smaller" activities.

The input audit data are divided into separate work streams according to the following logic. A workflow begins if:

- The first action has been executed on the current day;
- After the completion of the fulfillment condition of a previous workflow.

A workflow ends if:

- More than 15 minutes have passed since the last actions;
- It was the last action executed within the day.

Workflows are made up of individual work units of IS users, which, in turn, are defined with the following conditions. A work unit starts if:

- The first action has been executed on the current day;
- After the completion of the fulfillment condition of a previous work unit.

A work unit ends if:

- More than 5 minutes have passed since the last action (on the same IS object);
- It was the last action executed within the day.

A data set prepared for the approbation of the method served as an input for the method algorithm for handling these workflows and work units.

3.4 Approbation of the Method in Calculating Work Efficiency Metrics

The metrics defined in the field of software development are applied in the proposed method of evaluating the efficiency parameters of the users of IS (employees) working remotely:

- **Time spent on actions in an IS (Work unit / Workflow / Day Hours)**. It is possible to define continuous workflows of executed actions expressed in hours. It is also possible to analyze interruptions and durations of these.
- **Actions points**. It is possible to assign a weight to each action according to the benefit of the content of this action, which the execution of this action gives to the resulting IS work object (IS sub-part, file, etc.).
- **Velocity**. It is calculated as the sum of action points performed during a given workflow time. It shows the amount of work done within the workflow.
- **Productivity**. It is calculated by dividing the number of points accumulated within a workflow by the time spent in this workflow.
- **A specific set of actions (a pattern) performed by the user of IS**. It is possible to analyze the sequence of executed actions, duplicate actions, most / less used actions, unused actions, action patterns, etc.

3.5 IS User's Hours, Velocity, and Productivity Metrics

All related metrics can be defined as follows:

- **Work hours** – calculated as a sum of time spent within each particular work unit (its end time minus its start time);
- **Work velocity** – this metric is adapted from the software development methodology and foresees the assignment of points to each possible action in the IS. These points are defined by the problem domain experts and describe how "valuable" each particular action is for achieving the goals of work in a particular IS. In this case, a day's velocity is calculated as

a sum of all points of all performed actions in this IS on that particular day.

- **Work productivity** – the most important evaluation metric that allows one to judge how efficiently a user performed (e.g., on a particular day). It is calculated by dividing the day's work velocity by the day's work hours.

It is also possible to calculate according to values as a normalized percentage. In this case, hours are normalized relative to a normal work day amount (in the scope of this research it was considered to be 8 working hours). Velocity is normalized according to the individual maximums velocity of a user multiplied by according hours worked in a day and divided by the number of daily hours, which were achieved on the same day when the maximum day's velocity was achieved by this particular user. And, productivity percentage is calculated as the day's velocity multiplied by the maximum individual productivity and divided by the day's hours.

It is possible to define the maximum productivity value by taking into consideration the productivity variance.

The productivity variance of all employees in the experimental dataset is shown in Figure 6.

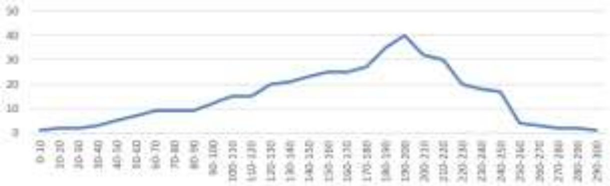


Fig. 6: Employees' productivity variance

The maximum expected employee productivity for employees of the same qualification can be defined as either the dispersion peak (200 points per hour) or, optionally – by increasing this peak value by 10%.

The relative values of hours, actions points, and productivity are shown in Figure 7, where the left column shows the dates of working days, and in the corresponding rows – the values of metrics of the particular user on a particular day in percentage against the maximum value of according to metric are being shown.

For example, the 1.67 hours worked by the user Abigail on May 4 has been converted to 27.78% in relation to 4.42 hours, which is the maximum number of hours worked by this employee in May (specifically – on the 20th of May). Here the amount of work means the hours spent working on the information system to perform the duties. Metric percentage values are colored red, orange, yellow, light green, and dark green, according to the following value ranges:

- red – (0% - 20%];
- orange – (20% - 40%];
- yellow – (40% - 60%];
- light green – (60% - 80%];
- dark green – (80% - 100%].

In the experimental data, the daily work efficiency metrics for users, Ava and Emily are also plotted in the view shown in Figure 8.

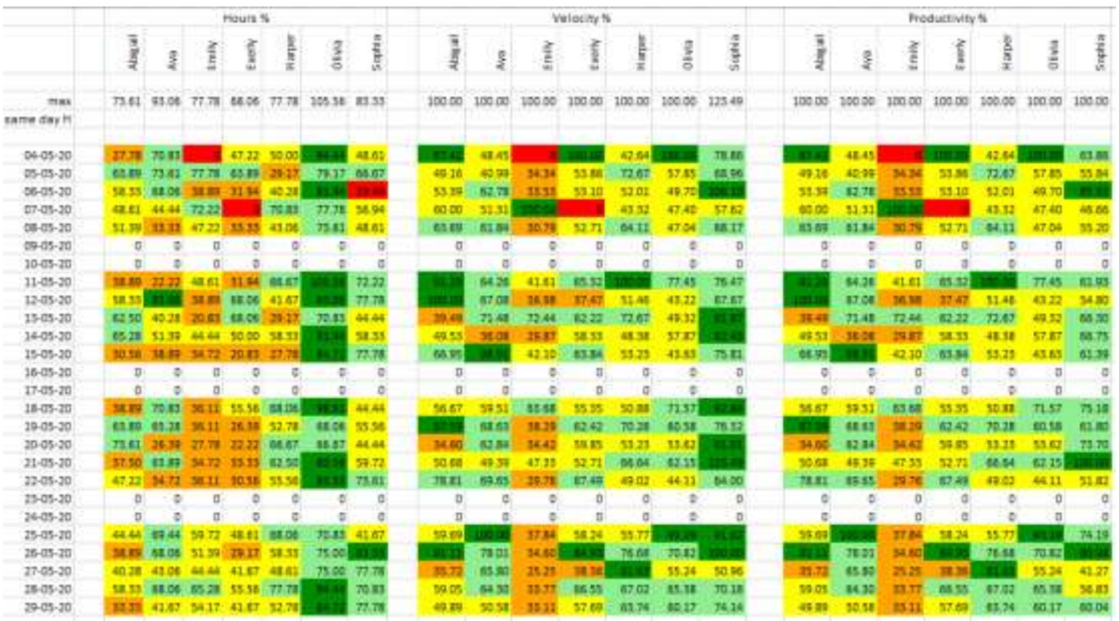


Fig. 7. Visualization of hours worked per day, velocity, and productivity percentages

	Hours %	Velocity %	Productivity %
	Abigail		
max	73.61	100.00	100.00
04.05.2020	27.78	87.42	87.42
05.05.2020	63.89	49.16	49.16
06.05.2020	58.33	53.39	53.39
07.05.2020	48.61	60.00	60.00
08.05.2020	51.39	63.69	63.69
09.05.2020	0	0	0
10.05.2020	0	0	0
11.05.2020	38.89	91.29	91.29
12.05.2020	58.33	100.00	100.00
13.05.2020	62.50	39.49	39.49
14.05.2020	65.28	49.53	49.53
15.05.2020	30.56	66.95	66.95
16.05.2020	0	0	0
17.05.2020	0	0	0

	Hours %	Velocity %	Productivity %
	Ava		
max	93.06	100.00	100.00
04.05.2020	70.83	48.45	48.45
05.05.2020	73.61	40.99	40.99
06.05.2020	68.06	62.78	62.78
07.05.2020	44.44	51.31	51.31
08.05.2020	33.33	61.84	61.84
09.05.2020	0	0	0
10.05.2020	0	0	0
11.05.2020	22.22	64.26	64.26
12.05.2020	93.06	67.08	67.08
13.05.2020	40.28	71.48	71.48
14.05.2020	51.39	36.08	36.08
15.05.2020	38.89	98.91	98.91
16.05.2020	0	0	0
17.05.2020	0	0	0

	Hours %	Velocity %	Productivity %
	Emily		
max	77.78	100.00	100.00
04.05.2020	0	0	0
05.05.2020	77.78	34.34	34.34
06.05.2020	38.89	33.53	33.53
07.05.2020	72.22	100.00	100.00
08.05.2020	47.22	30.79	30.79
09.05.2020	0	0	0
10.05.2020	0	0	0
11.05.2020	48.61	41.61	41.61
12.05.2020	38.89	36.98	36.98
13.05.2020	20.83	72.44	72.44
14.05.2020	44.44	29.87	29.87
15.05.2020	34.72	42.10	42.10
16.05.2020	0	0	0
17.05.2020	0	0	0

Fig. 8: Visualization of hours worked per day, percentages of performance and productivity, highlighting days worked in remote work mode

In addition, in the calendar, the days when the employee was working in a remote mode are highlighted in grey color.

Thus, the employee's "Ava" remote work was arranged on Wednesdays and the employee's "Emily" remote work was arranged on Mondays, Wednesdays, and Fridays. By analyzing the values of the metrics on different days, it can be concluded that the decreases and increases in work efficiency are dependent on the work mode, if any.

3.6 Metrics of IS User's Actions Patterns

One area of data mining that is also applicable to monitoring work efficiency is related to pattern recognition.

In the context of work efficiency, it is possible to define a set of actions performed by a specific IS user, analyze the sequence of executed actions / duplicate actions / the most popular / least used actions / unused actions/actions patterns, etc., and by comparing these sets between different users and users' groups.

The simplest case of action analysis in the context of their repetition is the analysis of the most popular actions within IS. In the case of an IS user, information about which actions are being used more often and which – less often is of particular interest to the problem environment expert.

This information is shown in Figure 9. Figure 9 shows only a list of the most popular actions performed by all users, as an example, which can then be compared to separate sets of actions performed by specific users, looking for habits of more / less / unused actions.

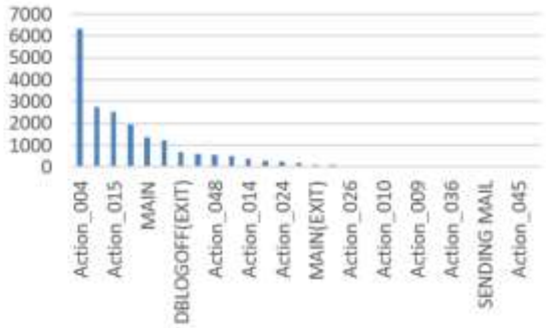


Fig. 9: Chart of the most popular actions of all users

This information can be useful for IS users' training purposes, for example – by noticing that one of the most "productive" actions is used less often by one particular IS user than other IS users, or, for example – that a less "desirable" action is used more often than others.

For a more detailed analysis of the behavior of users of IS–s, authors propose to use a combination of the Generalized Sequential Pattern (GSP), [10], algorithm, which is used to extract sequences, and a method of discovering sequential exceptional patterns using pattern growth, [11], by applying it to a set of collected users' actions.

In short, user sessions are defined by tuples $<A_1, A_2, \dots, A_i, \dots, A_n>$, where A_i is an action performed by an IS user.

Sets of all actions to be executed for each user are defined, and identical chains of action repetitions are found in these, for each user, chains of repetitions that occur more than 10 times in each of the action sets of the corresponding user. This results in more than 1000 patterns, each of which consists of 2–10 actions executed sequentially.

Thus, the example set of patterns for further analysis might look like this:

- *P1 = <A1, A2, A3>;*
- *P2 = <A3, A1, A3>;*
- *P3 = <A4, A2, A3, A4>;*
- *P4 = <A2, A1, A4, A5>;*
- *P5 = <A7, A7, A8>.*

The obtained dataset has been submitted for the approbation of the implementation of this metric, and, as a result, visualization of this data has been obtained in the form shown in Figure 10.

The problem domain expert analyses the patterns to determine which of these actions' chains are desirable or efficient to use, and which, on the contrary, are undesirable (e.g., such as opening a file, performing no actions, and then closing the file).

Thus, the resulting chains of repeated actions can be divided into patterns and anti-patterns.

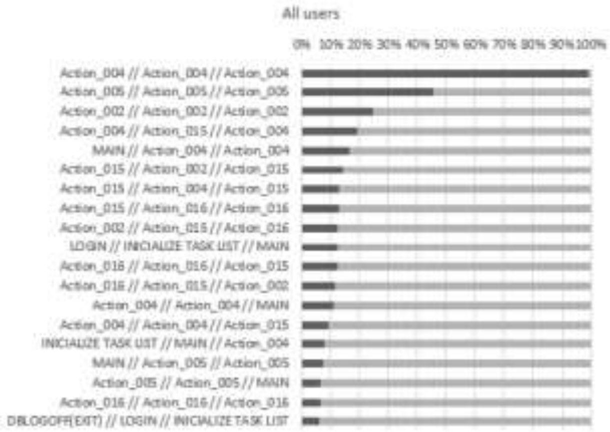


Fig. 10: Normalized number of patterns used by all users of IS

All users' sessions are submitted for analysis and patterns chains are searched for in users' actions. The number of patterns encountered by each user is noted, which is further normalized against the number of sessions and activities.

As a result, it is possible to define for each user a list of patterns/anti-patterns that he/she uses or does not use, and to conclude on the need for additional user training, or to detect less efficient/unwanted actions for further analysis and prevention of these.

Each pattern is defined as useful or useless based on the aggregate performance of the actions in it, divided by the number of actions in the pattern.

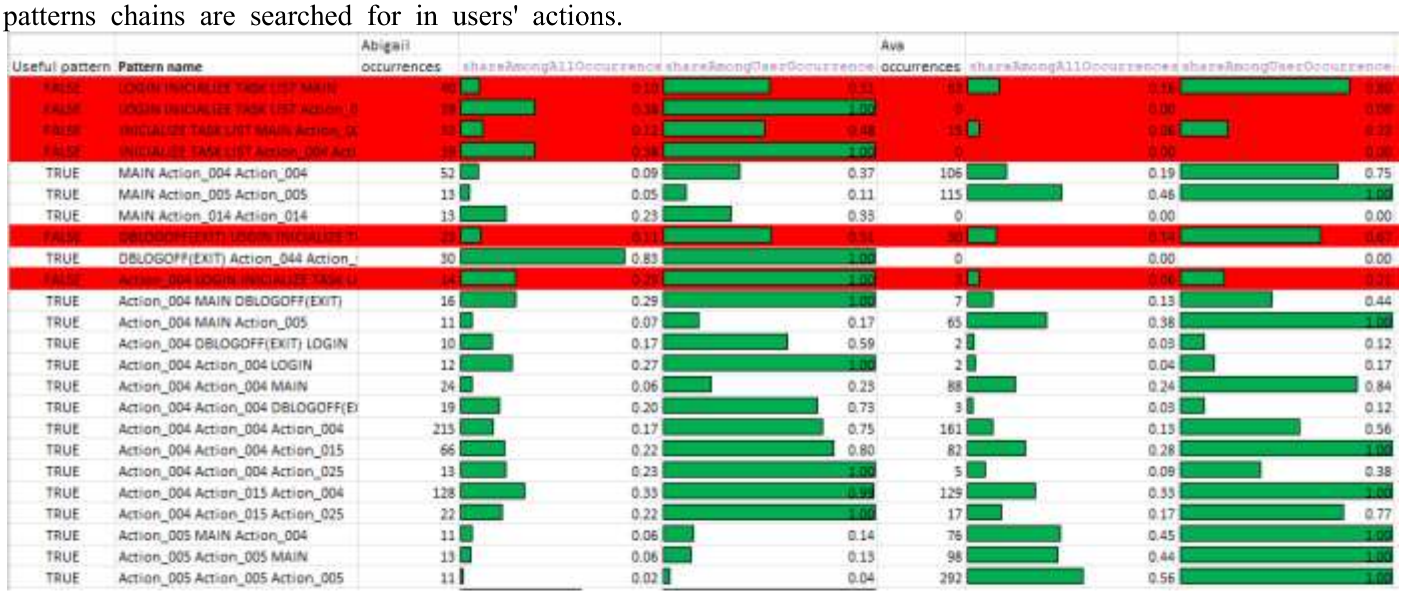
Two additional parameters for the characterization of patterns are calculated:

- The ratio of the number of uses of a certain pattern for a certain user to the total number of uses of the same pattern for all users.
- The ratio of the number of uses of a certain pattern for a certain user to the maximum number of uses of the same pattern for any of the users.

Patterns with absolute performance greater than *K* are considered efficient, while patterns with absolute performance less than *K* are considered less efficient. In real data, the value of *K* is determined when the data regarding the actions of real IS users is obtained.

In this study, this threshold *K* is assumed to be 5. A snippet of such information for further analysis is shown in Figure 11.

In this way, it is possible to define groups with similar user behavior to determine those users, who (according to the opinion of the problem domain expert), should belong to a certain group of users.



		Abigail			Ava		
Useful pattern	Pattern name	occurrences	shareAmongAllOccurrence	shareAmongUserOccurrence	occurrences	shareAmongAllOccurrences	shareAmongUserOccurrence
FALSE	LOGIN INICIALIZE TASK LIST MAIN	40	0.10	0.33	53	0.16	0.80
FALSE	LOGIN INICIALIZE TASK LIST Action_0	19	0.36	1.00	0	0.00	0.00
FALSE	INICIALIZE TASK LIST MAIN Action_0	33	0.12	0.48	15	0.06	0.22
FALSE	INICIALIZE TASK LIST Action_004 Act	19	0.36	1.00	0	0.00	0.00
TRUE	MAIN Action_004 Action_004	52	0.09	0.37	106	0.19	0.75
TRUE	MAIN Action_005 Action_005	13	0.05	0.11	115	0.46	1.00
TRUE	MAIN Action_014 Action_014	13	0.23	0.35	0	0.00	0.00
FALSE	DBLOGOFF(EXIT) LOGIN INICIALIZE T	23	0.11	0.31	30	0.14	0.67
TRUE	DBLOGOFF(EXIT) Action_044 Action_	30	0.83	1.00	0	0.00	0.00
FALSE	Action_004 LOGIN INICIALIZE TASK L	14	0.29	1.00	3	0.06	0.21
TRUE	Action_004 MAIN DBLOGOFF(EXIT)	16	0.29	1.00	7	0.13	0.44
TRUE	Action_004 MAIN Action_005	11	0.07	0.17	65	0.38	1.00
TRUE	Action_004 DBLOGOFF(EXIT) LOGIN	10	0.17	0.59	2	0.03	0.12
TRUE	Action_004 Action_004 LOGIN	12	0.27	1.00	2	0.04	0.17
TRUE	Action_004 Action_004 MAIN	24	0.06	0.23	88	0.24	0.84
TRUE	Action_004 Action_004 DBLOGOFF(E	19	0.20	0.73	3	0.03	0.12
TRUE	Action_004 Action_004 Action_004	215	0.17	0.75	161	0.13	0.56
TRUE	Action_004 Action_004 Action_015	66	0.22	0.80	82	0.28	1.00
TRUE	Action_004 Action_004 Action_025	13	0.23	1.00	5	0.09	0.38
TRUE	Action_004 Action_015 Action_004	128	0.33	0.99	129	0.33	1.00
TRUE	Action_004 Action_015 Action_025	22	0.22	1.00	17	0.17	0.77
TRUE	Action_005 MAIN Action_004	11	0.06	0.14	76	0.45	1.00
TRUE	Action_005 Action_005 MAIN	13	0.06	0.13	98	0.44	1.00
TRUE	Action_005 Action_005 Action_005	11	0.02	0.04	292	0.56	1.00

Fig. 11: Representation of patterns, highlighting efficient and less efficient actions combinations

The parameters characterizing the patterns are displayed in the columns of the table as bar graphs, which represent the obtained ratio values of the parameter.

For patterns with “Useful pattern” set to False, the row in the table is colored red and the bar in green represents the resulting pattern value for 1 (100%).

4 Conclusion

Analysis, evaluation, and monitoring of work efficiency is one of the most requested activities in a company of any type and field.

Evaluating work efficiency is one of the important tasks in company management and monitoring employee activities. Knowledge of how to solve it will positively affect the business environment and its development in various aspects.

This task is especially relevant if the company’s domain of activity is complicated, requires the use of several IS-s and work support tools, and the specifics and organization of the work allow employees to perform their duties both remotely and in the office.

The last case further accentuates the fact that solving such a task of accounting and efficiency assessment is not a trivial accumulation of statistical data, but requires more advanced methods instead.

In the course of this research, a solution has been developed, where the environment for determining the efficiency of employees is any kind of usage of one or more IS-s, within which the fulfillment of the employee’s work duties takes place.

This process can take place both remotely (isolated from the work environment and the collective of colleagues) and also – in office premises. Another aspect is related to the fact that an employee can work both with only one IS and with several IS-s in parallel, which requires the simultaneous integration of several data sources and the parallel reading and analysis of relevant data.

Also, it is potentially allowed that the metrics of employee efficiency parameters and calculations are compared with information from project management tools, in which the employee reports on the progress of his tasks.

In general, in the course of the research, the input data was defined, which allows evaluation of the efficiency of the company’s working employees with the proposed metrics, which are used within the developed method.

It was identified that the office-based / remote working employee is in daily contact with several specific / standard IS products and that the research results are universal and can be applied in any IT-related industry.

In this paper, the application of the solution is demonstrated on anonymized data obtained from a real company for a period of 10 months.

Therefore, the application of the solution developed in the course of this research will be suitable for any work environment, in which it will be possible to accumulate data on actions performed by employees and their execution time in various types of IS-s (for example – while work with clients and partners, documents processing, user support, digitization of any type of content, sales, and marketing, etc.).

Work is currently underway on further improvement of a method support tool. The developed solution could potentially be interesting for managers, quality specialists, and employees at all levels of the company operating in any problem domain.

References:

[1] Kittisak Umaji, Worapat Paireekreng, "A Study of the Remote Working Efficiency in IT Project Implementation during the COVID-19 Pandemic", WSEAS Transactions on Business and Economics, vol.20, pp. 400-409, 2023

[2] International Labour Organization, *“Practical guide on teleworking during the COVID-19 pandemic and beyond”*. [Online]. Available on: https://www.ilo.org/moscow/news/WCMS_751232/lang--en/index.htm

[3] Mahdani Ibrahim, Jumadil Saputra, Muhammad Adam, Mukhlis Yunus, "Organizational Culture, Employee Motivation, Workload and Employee Performance: A Mediating Role of Communication", WSEAS Transactions on Business and Economics, vol. 19, pp. 54-61, 2022

[4] Buffer, *“State Of Remote Work 2020”*, 2020. [Online]. Available on: https://buffer.com/state-of-remote-work/2020

[5] J. Andrews, *“20 tips for remote employee performance reviews”*, [Online]. Available on: https://performancereviews.net/8-tips-for-evaluating-telework

[6] O. Nikiforova, V. Zabiniako, J. Kornienko, P. Garkalns, R. Rizhko and M. Gasparoviča-Asīte, “Solution to CAD Designer Effort Estimation Based on Analogy with Software Development Metrics”, *In Proceedings of*

International Conference „Evaluation of Novel Approaches to Software Engineering" (ENASE 2022), pp. 1-9 – in press.

[7] O. Nikiforova, V. Zabiniako, J. Kornienko, R. Rizhko, K. Babris, V. Nikulsins, P. Garkalns, M. Gasparoviča-Asīte, "Solution to On-line vs On-site Work Efficiency Analysis on the Example of Engineering System Designer Work", *Applied Computer Systems*, vol.26, no.2, pp. 87-95, 2021.

[8] O. Nikiforova, V. Zabiniako, J. Kornienko, R. Rizhko, K. Babris, M. Gasparoviča-Asīte, "Efficiency Monitoring of Engineering System Designer Work Based on Multi-System User Behavior Analysis with AI/ML Algorithms", *2021 IEEE 62nd International Scientific Conference on Power and Electrical Engineering of Riga Technical University*, Riga, Latvia, 2021.

[9] O. Nikiforova, V. Zabiniako, P. Garkalns, J. Kornienko, I. Volodko, R. Rizhko "Definition of Metrics for Work Efficiency Monitoring Based on Multi-System Usage Behaviour Analysis", *In Proceedings of 17th Iberian Conference on Information Systems and Technologies*, pp. 1-6.

[10] R. Srikant, R. Agrawal, "Mining Sequential Patterns: Generalizations and Performance Improvements". *In Proceedings of the 5th International Conference on Extending Database Technology: Advances in Database Technology (EDBT '96)*, Peter M. G. Apers, Mokrane Bouzeghoub, and Georges Gardarin (Eds.). Springer-Verlag, London, UK, 3–17. 1996.

[11] D. Mollenhauer, M. Atzmueller, *"Sequential Exceptional Pattern Discovery Using Pattern-Growth: An Extensible Framework for Interpretable Machine Learning on Sequential Data"*, 2020. [Online]. Available on: http://ceur-ws.org/Vol-2796/xi-ml-2020_mollenhauer.pdf

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

-Oksana Ņikiforova introduced an idea to adapt metrics used for work estimation in software development for any problem domain, where it is possible to collect data about workers' activities and formulated all the calculations and dependencies for these metrics.
-Vitaly Zabiniako has implemented algorithms for metrics calculation and performed visualization of the data.
-Pavels Garkalns and Jurijs Kornienko have been working on the idea of pattern analysis for user behavior.
-Vladimirs Nikulsins and Andrejs Romanovs have been working on data simulation and optimization.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

The research is leading to Specific Objective 1.1.1 "Improve research and innovation capacity and the ability of Latvian research institutions to attract external funding, by investing in human capital and infrastructure" 1.1.1.1. measure "Industry-Driven Research" - Round 4. The project name is Analysis of Efficacy and Behavior of Remote Users of IT Systems Using AI/ML".

The research has been supported by the Doctoral Grant program of Riga Technical University.

Conflict of Interest

The authors have no conflict of interest to declare.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US