

PREDICTING TRAINING READINESS FROM WEARABLE-DEVICE DATA USING HEART RATE RECOVERY

Kantayeva D.

Astana IT University, Astana, Kazakhstan

Corresponding author: Kantayeva D. d.kantayeva@astanait.edu.kz

Abstract. Predicting an athlete's subjective training readiness from wearable-device data is a fast-growing field. Heart rate recovery (HRR) is the rate of heart-rate decline after exercise and a recovery indicator recognized in sports physiology that is not displayed to the user in the Fitbit app. The open PMData dataset holds 1,757 daily records from 16 recreational participants collected between November 2019 and March 2020. From the heart-rate data in this dataset, HRR60 was computed – to the best of our knowledge, for the first time for this dataset – after each of 772 training sessions (495 with a valid result), and it was tested whether it adds value to readiness prediction beyond metrics already visible to the user, using Ridge, Random Forest, and Gradient Boosting regression with Leave-One-Participant-Out validation. Neither the ablation test, nor permutation feature importance, nor direct correlation with same-day or next-day readiness showed a meaningful gain in accuracy. This result tempers overstated expectations of "hidden" wearable-device metrics and points to the need for ablation testing of such hypotheses.

Key words: heart rate recovery; training readiness; wearable devices; machine learning; training load.

Introduction. Wearable devices such as Fitbit and others have made continuous physiological data collection accessible to recreational athletes. These devices provide the user with metrics such as sleep score and resting heart rate, but many physiological indicators are neither computed nor displayed to the user by the app. The accuracy of consumer heart-rate measurements requires separate methodological scrutiny [1]. Reviews of HR-based methods for monitoring athletes' training status note that inconsistency in results stems from methodological differences between studies rather than from limitations of the measurements themselves [2]. Recent validation studies further show that the accuracy of consumer photoplethysmography-based heart-rate sensors varies by device and by exercise intensity, remaining imperfect during dynamic movement even as it improves at higher effort levels [3, 4].

One such indicator is heart rate recovery (HRR), which reflects the rate of heart-rate decline after the cessation of physical exertion. A literature review confirms HRR's status as a recognized indicator of recovery and training status [5]. However, in elite runners the development of overreaching was accompanied by a faster, rather than slower, HRR [6], which points to the complexity of this indicator and to the need for empirical testing of its predictive value. Machine-learning models built on broader, multi-feature physiological and training-load profiles have recently been shown to predict daily recovery status in endurance athletes with reasonable accuracy [7], which motivates the narrower question addressed here: whether HRR specifically, rather than a wearable-derived feature set in general, adds value beyond metrics already available to the user.

A study of recreational athletes' motivation during competition preparation on a Kazakhstani sample [8] shows that recreational athletes are motivated in training more by health improvement and self-development than by results themselves. It is therefore important for this group to understand whether the "hidden" metrics of their wearable devices genuinely carry additional useful information, or whether this is merely a marketing expectation.

The aim of this work is an empirical test of whether heart rate recovery, computed from the PMData dataset [9], adds predictive value to a model of daily training readiness beyond metrics

already visible to the user in Fitbit. Standard regression models are used, and the scientific novelty lies in the data: to the best of our knowledge, HRR60 is computed from heart-rate data for the first time for this dataset, and its value is assessed quantitatively, rather than declaratively, through an ablation test, permutation feature importance, and direct correlation with readiness.

Materials and Methods. This study is a secondary analysis of the open PMData dataset [9] (Simula Research Laboratory, Norway, CC BY 4.0 license): 16 physically active recreational adult participants, from November 2019 to March 2020, daily completed a wellness questionnaire, including readiness on a 0–10 scale, and a training report including type, duration, and rating of perceived exertion (RPE) for each of 772 sessions, while continuously wearing a Fitbit tracker [9].

Heart rate recovery calculation. For each training session, the start and end times were determined. Points with confidence ≥ 2 (medium and high signal quality) were selected, and the session peak heart rate (hr_peak) was determined as the maximum among points within the interval from the start to the end of the session. Recovery heart rate at 60 seconds after the end of the session was determined using the nearest-in-time point with a tolerance of 30 seconds; if no suitable point was found, the value was treated as missing. The indicator $HRR60 = hr_peak - hr(\text{end of session} + 60 \text{ s})$. A valid HRR60 value could be computed for 495 of the 772 training sessions, or 64.1%, across 15 of the 16 participants; the mean HRR60 value was 45.8 bpm, with a standard deviation of 27.9 and a range from -25 to 119 bpm. This 60-second window follows the general logic of standard heart-rate-recovery protocols, which define HRR as the heart-rate drop over a fixed interval after peak effort [5]; unlike those protocols, which typically use a controlled, supervised recovery period, HRR60 here is derived from real-world training sessions with a self-reported end time and uncontrolled post-session behaviour, and the underlying photoplethysmography signal is known to be less accurate during and immediately after vigorous, dynamic exercise [3].

Feature engineering. The unit of analysis is a day-observation within a participant; the target variable is the readiness value on a 0–10 scale on day t (regression). All features for date t are constructed exclusively from data up to and including day t minus one (lag $t-1$ and earlier). The base set of features includes: training load on $t-1$ and its rolling sums, including the acute:chronic workload ratio, ACWR [10]; wellness on $t-1$ (readiness, fatigue, soreness, stress, mood) and their 3- and 7-day rolling averages; Fitbit physiology on $t-1$ (sleep score, resting heart rate); day of week and the sequential day-observation number. The heart rate recovery feature is the mean HRR60 across training sessions on day $t-1$ and a binary indicator of whether at least one session with a valid HRR value occurred that day (has_hrr_t1). The final dataset, after excluding observations without readiness($t-1$), comprised 1,423 day-observations across 16 participants.

Data quality and measurement limitations. A valid HRR60 could not be computed for 277 of the 772 sessions (35.9%): most commonly because no heart-rate sample met the confidence ≥ 2 threshold within the 60 ± 30 -second recovery window, which typically occurs when the device is removed immediately after exercise or when motion artefacts degrade the photoplethysmography signal during the final, most intense phase of a session [3]. These sessions were treated as missing at the session level and were not imputed; days with zero valid HRR sessions likewise received a missing value for the daily feature $hrr_60_mean_t1$. For model fitting only, this missing value was imputed with the sample median, and the accompanying binary indicator has_hrr_t1 (0 for imputed days) was retained as a separate feature so that the models could distinguish a genuinely low recovery value from an absent measurement. A small fraction of valid sessions (13 of 495, 2.6%) showed a negative HRR60, and the observed range extended up to 119 bpm; both patterns are plausible under an uncontrolled field protocol — a negative value can arise when the true peak was underestimated due to a brief signal dropout at the point of maximal exertion, or when heart rate continued to rise briefly after the recorded session-end time, while an unusually large value can arise when the 30-second tolerance window captures a reading close to 90 seconds after session end, allowing more actual recovery time than the nominal 60 seconds. None of these values were excluded as outliers, since no external reference measurement was available to distinguish genuine physiological variability from measurement noise.

Models, validation, and testing the value of HRR. Three regression models were used:

1. Ridge regression (alpha = 5, with feature standardization);
2. Random Forest Regressor (300 trees, maximum depth 6, minimum 5 samples per leaf);
3. Gradient Boosting Regressor (200 trees, maximum depth 2, learning rate 0.05).

Model quality was assessed via Leave-One-Participant-Out cross-validation, in which no participant appears in the training and test sets simultaneously. The value of HRR was tested via an ablation test: each model was trained in two feature variants — with HRR (HRR60 on t-1 and has_hrr_t1) and without it — with all other features held equal. The quality metrics were mean absolute error (MAE) and root mean squared error (RMSE) on the readiness scale (0–10); the coefficient of determination R^2 is reported as a secondary metric. Permutation feature importance (30 repeats, based on MAE) was additionally computed for the Gradient Boosting model trained on the full dataset, along with the direct Pearson correlation between HRR60 and readiness. The analysis was performed in Python using the pandas, scikit-learn, and scipy libraries.

Results. The final modeling dataset comprises 1,423 day-observations across 16 participants. The target variable, readiness on day t, had a mean of 4.91 (standard deviation 1.92) on the 0–10 scale; the lagged readiness feature used as a predictor, readiness(t-1), was nearly identical (mean 4.93, SD 1.90), reflecting the strong day-to-day autocorrelation of this self-reported measure. The ablation test (Table 1) shows that adding HRR features to the base feature set produces only minimal changes in the metrics: MAE decreases by no more than 0.015 (Gradient Boosting), and R^2 increases by no more than 0.011 — differences that do not exceed the noise level given the pilot sample size; for Random Forest the difference is practically absent ($R^2 = 0.422$ in both variants).

Table 1 — Ablation test: model with and without HRR features (Leave-One-Participant-Out)

Model	Feature variant (HRR/-)	MAE	RMSE	R^2
Ridge	-	1,075	1,419	0,455
Ridge	HRR	1,074	1,417	0,457
Random Forest	-	1,114	1,462	0,422
Random Forest	HRR	1,111	1,461	0,422
Gradient Boosting	-	1,116	1,463	0,421
Gradient Boosting	HRR	1,101	1,449	0,432

Compiled by the author based on an analysis of PMData [9].

The overall distribution of HRR60 across the 495 sessions is shown in Figure 1: values range from -25 to 119 bpm, with a mean of 45.8 and a median of 45.0.

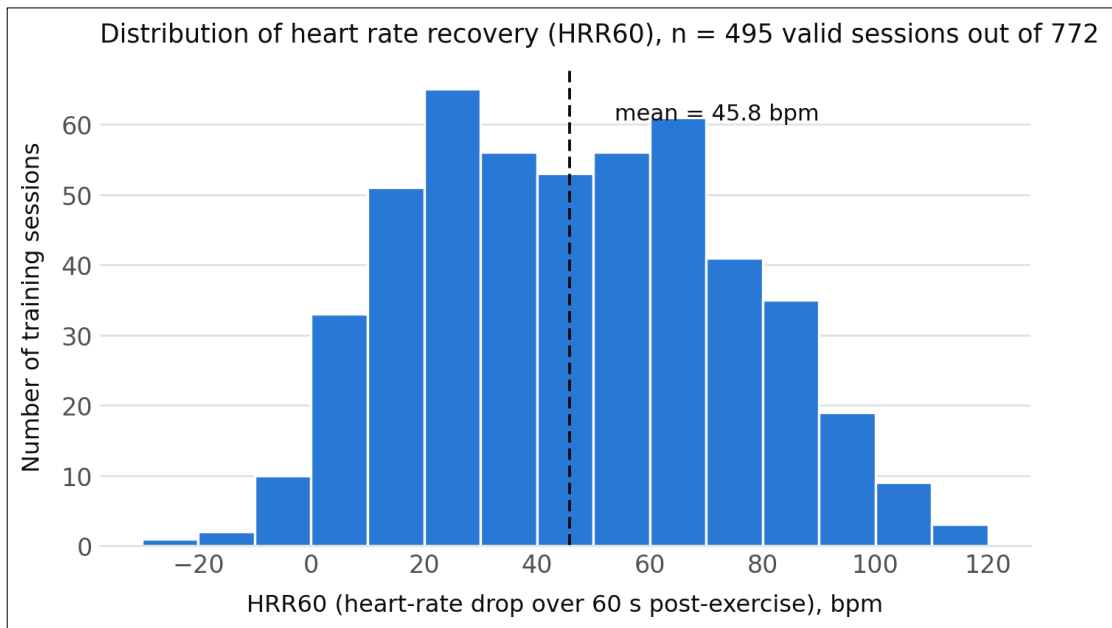


Figure 1 — Distribution of heart rate recovery (HRR60) across 495 training sessions

Compiled by the author based on an analysis of PMData [9].

A separate test of the relationship between HRR60 and readiness revealed no correlations: between mean daily HRR60 and same-day readiness, $r = -0.01$ ($p = 0.86$); between day- t HRR60 and day- $(t+1)$ readiness, $r = -0.07$ ($p = 0.15$, $n = 406$). Figure 2 visually confirms the absence of a relationship.

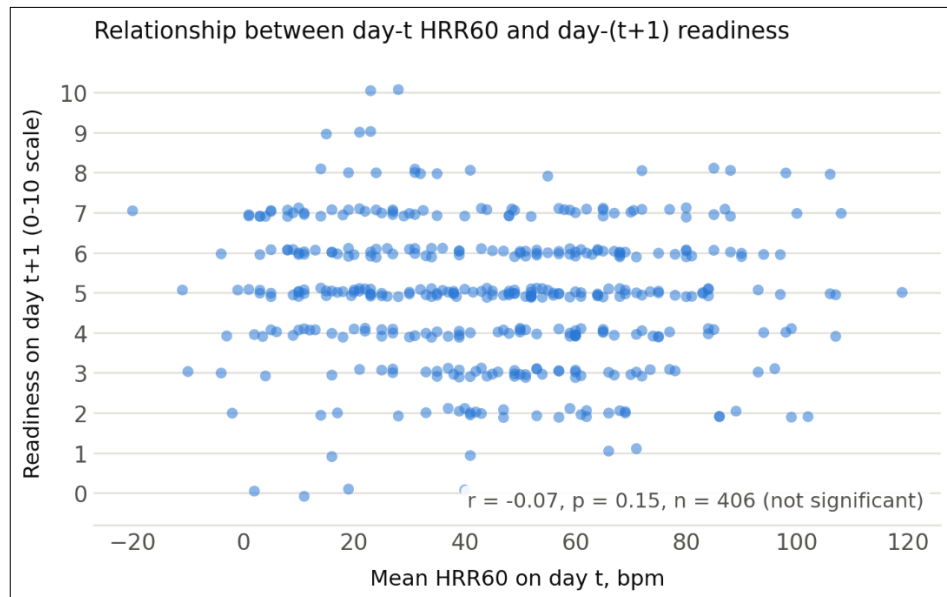


Figure 2 — Relationship between mean day- t HRR60 and day- $(t+1)$ readiness

Compiled by the author based on an analysis of PMData [9].

Permutation feature importance shows that, of 24 features, HRR60 (`hrr_60_mean_t1`) ranks only 18th by contribution to error reduction, while the binary indicator of a valid HRR measurement on the previous day (`has_hrr_t1`) ranks 4th, noticeably ahead of the HRR value itself. Thus, the ablation test, permutation importance, and direct correlation all point to the absence of any meaningful predictive value of HRR60 in this task.

Discussion. The main result of this work is an honest null result. HRR does not add predictive value beyond data already visible to the user in Fitbit, whether by the ablation test, by permutation feature importance, or by correlation with readiness: $r = -0.01$ and $r = -0.07$, both non-significant. This conclusion is necessarily bounded by the conditions of the present study – a single, pilot-scale cohort, the specific HRR60 computation protocol used here, and the particular feature set and models tested – and should not be read as a general claim that heart-rate recovery has no value for training-readiness monitoring under other measurement protocols or study designs.

Possible reasons fall into several categories. The relationship between HRR and the body's functional state is nonlinear — in elite athletes, the development of overreaching is accompanied by a faster, rather than slower, HRR [6], and protocols for measuring and interpreting HRR differ noticeably between studies [2, 5]. Wearable devices are susceptible to motion artifacts immediately after exertion [1], and recent validation studies confirm that consumer photoplethysmography accuracy remains imperfect during and shortly after vigorous exercise even for the best-performing devices [3, 4]. The end time of a training session here was taken from self-report rather than from the Fitbit exercise log, and a valid HRR60 value was obtained for only 64.1% of sessions for the reasons described above. A day-averaged value is also used, whereas individual sessions within a day can differ in intensity.

It is worth noting that the broader machine-learning literature on recovery prediction from wearable and training-load data has reported more encouraging results using richer, multi-feature profiles [7]; the null result obtained here therefore appears specific to HRR60 as a single feature under the present measurement protocol, rather than evidence that wearable-derived recovery indicators cannot contribute to readiness prediction in general.

Notably, the presence of a valid measurement on the previous day (`has_hrr_t1`) proved more informative, ranking 4th in importance versus 18th for `hrr_60_mean_t1`. It likely reflects, indirectly, the fact and nature of the training load rather than the quality of physiological recovery itself.

The result is practically relevant for recreational athletes [8]: the test tempers overstated expectations of “hidden” wearable-device metrics and is consistent with broader calls in the wearable-biosensing and sports-technology literature for rigorous, feature-level validation before such metrics are built into athlete-facing training and coaching tools [11]. Limitations of this study include: a sample of 16 participants (LOPO validation), a self-reported target variable, permutation importance used instead of SHAP, a dataset collected in Norway in 2019–2020, and associative rather than causal relationships.

Conclusion. This work computed the heart rate recovery indicator (HRR60) from heart-rate data after each training session for the PMData dataset – to the best of our knowledge, for the first time for this dataset – and conducted an ablation test of the hypothesis that it adds value for predicting a recreational athlete's training readiness. Within the boundaries of this sample, this HRR60 computation protocol, and the models and features tested, adding HRR did not improve prediction accuracy, either by MAE/RMSE in the ablation test, by permutation feature importance, or by direct correlation with same-day or next-day readiness. This null result, specific to the present conditions, underscores the need for ablation testing rather than declarative claims of value.

Directions for future work include: modeling HRR at the level of individual sessions rather than a daily aggregate; using a more precise source for training end time to reduce missing values; increasing and diversifying the sample; external validation on an independent sample; and comparing HRR with other activity markers (for example, heart rate variability) as potentially more informative candidates for the role of a “hidden” readiness metric.

ПРОГНОЗИРОВАНИЕ ГОТОВНОСТИ К ТРЕНИРОВКЕ ПО ДАННЫМ НОСИМЫХ УСТРОЙСТВ С УЧЕТОМ ВОССТАНОВЛЕНИЯ ЧАСТОТЫ СЕРДЕЧНЫХ СОКРАЩЕНИЙ

Кантаева Д.

Astana IT University, г. Астана, Казахстан

Автор для корреспонденции: Кантаева Д. d.kantayeva@astanait.edu.kz

Аннотация. Прогнозирование субъективной готовности спортсмена к тренировке по данным носимых устройств является быстро растущей областью. Heart rate recovery (HRR) это скорость снижения пульса после нагрузки, признанный в спортивной физиологии индикатор восстановления, который не отображается пользователю в приложении Fitbit. В открытой базе данных PMData хранятся 1757 дневных записей 16 участников-любителей в период с ноября 2019 года по март 2020. Из показателей пульса этих данных, насколько нам известно, впервые для этого датасета вычислен HRR60 после каждой из 772 тренировочных сессий (495 – с валидным результатом), а также проверено, добавляет ли он ценность прогнозу готовности сверх видимых пользователю метрик в регрессии Ridge, Random Forest, Gradient Boosting и Leave-One-Participant-Out валидации. Ни абляционный тест, ни перестановочная важность признаков, ни прямая корреляция с готовностью того же или следующего дня не показали заметного прироста точности. Это результат, который ограничивает завышенные ожидания от «скрытых» метрик носимых устройств и указывает на необходимость абляционной проверки таких гипотез.

Ключевые слова: heart rate recovery; готовность к тренировке; носимые устройства; машинное обучение; тренировочная нагрузка.

ЖҮРЕК СОҒУ ЖИІЛІГІНІҢ ҚАЛПЫНА КЕЛУІН ЕСКЕРУ АРҚЫЛЫ КИІЛЕТІН ҚҰРЫЛҒЫЛАР ДЕРЕКТЕРІ БОЙЫНША ЖАТТЫҒУҒА ДАЙЫНДЫҚТЫ БОЛЖАУ

Кантаева Д.

Astana IT University, Астана қ., Қазақстан

Хат алмасу үшін автор: Кантаева Д. d.kantayeva@astanait.edu.kz

Аңдатпа. Спортшының жаттығуға субъективті дайындығын киілетін құрылғылар деректері бойынша болжау жедел дамып келе жатқан сала. Жүрек соғу жиілігінің қалпына келуі (HRR) жүктемеден кейінгі пульстің төмендеу жылдамдығы, спорт физиологиясында танылған қалпына келу көрсеткіші, ол Fitbit қосымшасында пайдаланушыға көрсетілмейді. PMData ашық дерекқорында 2019 жылғы қарашадан 2020 жылғы наурызға дейінгі кезеңде 16 әуесқой қатысушының 1757 күндік жазбасы сақталған. Осы деректердегі пульс көрсеткіштерінен бұл деректер жиынтығы үшін, біздің білуімізше, алғаш рет 772 жаттығу сеансының (495-і жарамды нәтижемен) әрқайсысынан кейін HRR60 есептелді, сондай-ақ ол Ridge, Random Forest, Gradient Boosting регрессиясы мен Leave-One-Participant-Out валидациясында пайдаланушыға бұрыннан көрінетін көрсеткіштерден тыс дайындық болжамына құндылық қосатыны тексерілді. Абляциялық тест те, белгілердің ауыстыру маңыздылығы да, сол күнгі немесе келесі күнгі дайындықпен тікелей корреляция да сезілерлік дәлдік өсімін көрсеткен жоқ. Бұл киілетін құрылғылардың «жасырын» көрсеткіштеріне деген асыра үміттенуді тежейтін нәтиже және мұндай гипотезаларды абляциялық тексеруден өткізу қажеттігін көрсетеді.

Түйін сөздер: heart rate recovery; жаттығуға дайындық; киілетін құрылғылар; машиналық оқыту; жаттығу жүктемесі.

References

- 1 Nelson B.W., Low C.A., Jacobson N., Areán P., Torous J., Allen N.B. Guidelines for wrist-worn consumer wearable assessment of heart rate in biobehavioral research // *npj Digital Medicine*. – 2020. – Vol. 3. – Art. 90. – DOI: 10.1038/s41746-020-0297-4.
- 2 Buchheit M. Monitoring training status with HR measures: do all roads lead to Rome? // *Frontiers in Physiology*. – 2014. – Vol. 5. – Art. 73. – DOI: 10.3389/fphys.2014.00073.
- 3 Jeong Y., Yoon S., Lee J., Jeong W.-M., Ha M.-S. Comparison of heart rate measurement accuracy among commercially available photoplethysmography-based wearable devices across exercise intensities // *Physical Activity and Nutrition*. – 2026. – Vol. 30, No. 2. – P. 142–150. – DOI: 10.20463/pan.2026.0032.
- 4 Lambe R., Baldwin M., O'Grady B., Schumann M., Caulfield B., Doherty C. The accuracy of Apple Watch measurements: a living systematic review and meta-analysis // *npj Digital Medicine*. – 2026. – Vol. 9. – Art. 63. – DOI: 10.1038/s41746-025-02238-1.
- 5 Daanen H.A.M., Lamberts R.P., Kallen V.L., Jin A., Van Meeteren N.L.U. A systematic review on heart-rate recovery to monitor changes in training status in athletes // *International Journal of Sports Physiology and Performance*. – 2012. – Vol. 7. – P. 251–260. – DOI: 10.1123/ijsp.7.3.251.
- 6 Aubry A., Hausswirth C., Louis J., Coutts A.J., Buchheit M., Le Meur Y. The development of functional overreaching is associated with a faster heart rate recovery in endurance athletes // *PLOS ONE*. – 2015. – Vol. 10, No. 10. – Art. e0139754. – DOI: 10.1371/journal.pone.0139754.
- 7 Rothschild J.A., Stewart T., Kilding A.E., Plews D.J. Predicting daily recovery during long-term endurance training using machine learning analysis // *European Journal of Applied Physiology*. – 2024. – Vol. 124, No. 11. – P. 3279–3290. – DOI: 10.1007/s00421-024-05530-2.
- 8 Krasnik Yu.N. Osobennosti sportivnoi motivatsii v period podgotovki k sorevnovaniyam u sportsmenov-professionalov i sportsmenov-lyubitelei [Features of sports motivation during competition preparation among professional and amateur athletes] // *Vestnik KazNU. Seriya psikhologii i sotsiologii*. – 2022. – Vol. 81, No. 2. – P. 60–68. – DOI: 10.26577/JPsS.2022.v81.i2.07.
- 9 Thambawita V., Hicks S., Borgli H. et al. PMData: a sports logging dataset // *Proceedings of the 11th ACM Multimedia Systems Conference (MMSys '20)*. – New York: ACM, 2020. – P. 231–236. – DOI: 10.1145/3339825.3394926.
- 10 Gabbett T.J. The training-injury prevention paradox: should athletes be training smarter and harder? // *British Journal of Sports Medicine*. – 2016. – Vol. 50, No. 5. – P. 273–280. – DOI: 10.1136/bjsports-2015-095788.
- 11 Madrigal-Cerezo R., Domínguez-Sanz N., Martín-Rodríguez A. Wearable Biosensing and Machine Learning for Data-Driven Training and Coaching Support // *Biosensors*. – 2026. – Vol. 16, No. 2. – Art. 97. – DOI: 10.3390/bios16020097.

Авторлар туралы ақпарат // Информация об авторах // Information about the Authors

Кантаева Дина – докторант (PhD), Astana IT University, г. Астана, Казахстан.

Кантаева Дина – PhD докторант, Astana IT University, Астана қ., Қазақстан.

Dina Kantayeva – PhD student, Astana IT University, Astana, Kazakhstan.

e-mail: d.kantayeva@astanait.edu.kz

ORCID iD: 0000-0002-4097-5078