REPORT
on the classification of activities using machine learning algorithms

Task: According to the data of the accelerometer from the mobile phone, it is necessary to classify what kind of activity a person is engaged in walking, standing, running, or climbing stairs.
Use SVM and random forest algorithms from the scikit-learn library. As characteristics, you can take indicators from the accelerometer in order to improve the performance of the algorithms, you can first prepare the dataset and calculate the time domain features.
Compare the results of both algorithms and different models with each other. Use the classification report method for comparison.

Within the framework of this study, the classification of human activities (walking, standing, running, climbing stairs) was carried out using machine learning algorithms, in particular, the support vector method (SVM) and random forest (Random Forest).

Initial data: Data from the mobile phone accelerometer were used for the classification. Each record contains acceleration vectors in three dimensions (X, Y, Z) for each of the considered activities.

Process of data preparation: In the process of data preparation for classification, time domain features were calculated, including statistical indicators such as mean, variance, standard deviation (std), root_mean_square, minimum and maximum value, and median. Additionally, for each measurement of acceleration, fast Fourier transforms (FFT) were used to obtain frequency characteristics: power, energy, magnitude, the area under the curve (area), maximum and minimum magnitude (max_magnitude, min_magnitude), maximum and minimum magnitude index (max_index, min_index), entropy, skewness, kurtosis, Interquartile range, mean absolute deviation.
These features were used to improve the quality of the classification and allow the algorithms to better distinguish between different activities. The results after using these features indicate their effectiveness in improving the accuracy and generalization ability of data models

Models: Two classification models were used: SVM and Random Forest. SVM uses a hyperplane to separate classes in the feature space, while Random Forest is an ensemble method that uses multiple decision trees to make a decision.

Results before the creation of new features:

SVM to create new features:
Accuracy: 0.89
Weighted F1-score: 0.88

	precision	recall	f1-score	support
idle	0.96	0.99	0.97	9306
running	0.93	0.90	0.92	30609
stairs	1.00	0.00	0.01	1537
walking	0.80	0.91	0.85	16706
accuracy			0.89	58158
macro avg	0.92	0.70	0.69	58158
weighted avg	0.90	0.89	0.88	58158

```
[[ 9185    85     0    36]
 [  362 27609     0  2638]
 [   13   372     4  1148]
 [   54  1531     0 15121]]
```

Random Forest before creating new features:
Accuracy: 0.93
Weighted F1-score: 0.92

	precision	recall	f1-score	support
idle	0.99	0.99	0.99	9306
running	0.95	0.95	0.95	30609
stairs	1.00	0.10	0.18	1537
walking	0.86	0.95	0.90	16706
accuracy			0.93	58158
macro avg	0.95	0.75	0.76	58158
weighted avg	0.94	0.93	0.92	58158

```
[[ 9242    56     0     8]
 [   67 29034     0  1508]
 [    4   428   150   955]
 [   19   893     0 15794]]
```

Results after creating new features:

SVM with selected new features:
Accuracy: 0.98
Weighted F1-score: 0.97

	precision	recall	f1-score	support
idle	1.00	1.00	1.00	220
running	1.00	1.00	1.00	689
stairs	1.00	0.17	0.29	30
walking	0.93	1.00	0.97	354
accuracy			0.98	1293
macro avg	0.98	0.79	0.81	1293
weighted avg	0.98	0.98	0.97	1293

```
[[220   0   0   0]
 [  0 689   0   0]
 [  0   0   5  25]
 [  0   0   0 354]]
```

Random Forest with selected new features:
Accuracy: 0.99
Weighted F1-score: 0.99

	precision	recall	f1-score	support
idle	1.00	1.00	1.00	220
running	1.00	1.00	1.00	689
stairs	0.85	0.77	0.81	30
walking	0.98	0.99	0.98	354
accuracy			0.99	1293
macro avg	0.96	0.94	0.95	1293
weighted avg	0.99	0.99	0.99	1293

```
[[220   0   0   0]
 [  0 689   0   0]
 [  0   0  23   7]
 [  0   0   4 350]]
```

Conclusions: The calculation of time domain features contributed to a significant improvement in the quality of classification for both algorithms - it allowed the algorithms to better distinguish between classes and increase accuracy. This can be explained by the fact that the new features provide the models with additional information about the nature of movements, which improves their ability to distinguish between different activities.
Furthermore, Random Forest outperformed SVM in the classification of this data after using the new features. This is probably because Random Forest can effectively use many features, and new features provide it with additional information to make more accurate decisions.