The market-leading companies desire to exploit strength of machine learning techniques to extract hidden, valuable knowledge from the huge, fast growing volumes of data. One of the most promising directions of that research is classification task, which is widely used in computer security, medicine, finance, or trade. Designing such solutions we should take into consideration that in the modern world the most of the data arrive continuously and it causes that smart analytic tools should respect this nature and be able to interpret so-called data streams. Unfortunatelly, such a sysytem should be ready to combat at least two phenomena: skewed class distribution and *concept drift*. Only a few of the authors distinguish the differences between imbalanced data stream classification problem and a scenario where the prior knowledge about the entire data set is given. This discrepancy is a result of the lack of the knowledge about the class distribution and this issue is notably present in the in the initial stages of data stream classification. Another difficulty is the presence of the phenomenon called (*concept drift*), what can usually lead to the classifier quality deterioration. The *concept drift* may have different nature, but it causes the change of the of probability characteristics of the decision task, e.g., it could lead to a change of the prior probabilities, i.e., the frequency at which the objects appear in the examined classes. A typical example of a such case is technical diagnosis in which the fault probability increases with utilization time and it may be a result of material fatigue.

The project "*Imbalanced data streams classification algorithms*"is an attempt to connect this two important research trends and proposes the following research hypothesis:

Data stream classifiers trained on the basis of learning methods taking into consideration data imbalance can outperform classifiers trained on the basis of algorithms which do not take this characteristic into consideration.

Conducted literature study indicates the need to develop imbalanced data stream classification methods with special attention to:

- Methods for determining the imbalance ratio since most of the existing algorithms assume its prior knowledge.
- Non-stationary data stream classification methods due to most works assuming the stationarity of the streams, ignoring the *concept drift*.
- Reducing the memory complexity of imbalanced data stream classification model due to some approaches assuming that the entirety of the forthcoming minority class objects are stored in memory.
- Dedicated methods for data pre-processing.
- Non-stationary data streams classification methods that are not based on the paradigm of classifier ensemble.
- Methods using active learning for imbalanced data stream classifiers.

The project will focus on the following research tasks:

- Developing classifiers for learning from stationary and nonstationary imbalanced data streams.
- Creating an open-source software library for imbalanced data stream classification.
- Evaluation of the proposed classification algorithms.

The developed methods could be used in medicine, because medical screening for a condition is usually performed on a large population of people without the condition, to detect a small minority with it (e.g., HIV prevalence in the USA is ca. 0.4%), banking to detect fraudent transactions, fault diagnosis to enumerate only a few.