speed. Until very recently there was little published research about successful early threat detection models.
Other authors proposed an ensemble of Machine Learning models as a probable way of solving abovementioned early detection problems. Therefore, in this work, authors perform an investigation of selected method ensembles and present results of the comparison.

Application of Machine Learning for MWE Identification

I. Bumbulienė[1], J. Mandravickaitė[1, 2], T. Krilavičius[1, 2]

[1] Baltic Institute of Advanced Technology
[2] Vytautas Magnus university
t.krilavicius@bpti.lt

Identification of Multiword Expressions is an important problem in Natural Language Processing, especially for machine translation and other semantic analysis tasks. Often, lexical association measures (LAM), such as pointwise mutual information (PMI), log likelihood ratio (LLR), Dice are used to identify MWE's. However, just LAMs are insufficient for MWE detection, especially for Lithuanian language, but could be very useful as additional features for Machine Learning (ML) algorithms. Early experiments with Lithuanian and Latvian languages show that using Random Forest with Resample filter, we can achieve almost 99% precision, 58% recall and 73% F-score.
We discuss experiments with delfi.lt based corpora, different features, including LAMs, as well as experiments with different ML methods, i.e., Naive Bayes, Random Forests, Support Vector Machines, Artificial Neural Networks and others.