# Application of Machine Learning in Feature Selection for Continuous Authentication

Jnana Deepika Bala, Advisor: Dr. Ambareen Siraj, Tennessee Tech

## Abstract

Single time or static authentication can be a potential security vulnerability for highly secure systems. Although there exists advanced authentication methods such as graphical passwords, one time passwords and biometric authentication, they still face the same problem of single time authentication where masqueraders can potentially hijack sessions once authorized users are logged in. To address this problem, continuous authentication monitors the user's activities from login to logout session. Machine learning can be used to aid continuous authentication given the ability to learn useful knowledge about authorized user without direct programming. Rather than manually selecting features to feed machine learning algorithms, feature selection algorithms can be used to yield better results.

In this project, we have experimented with various combinations of feature selection and machine learning algorithms to learn models of authorized user keystroke dynamics more efficiently during the authentication process. Preliminary results show that such additional levels of data analytics helps to improve unauthorized user detection.

## Background

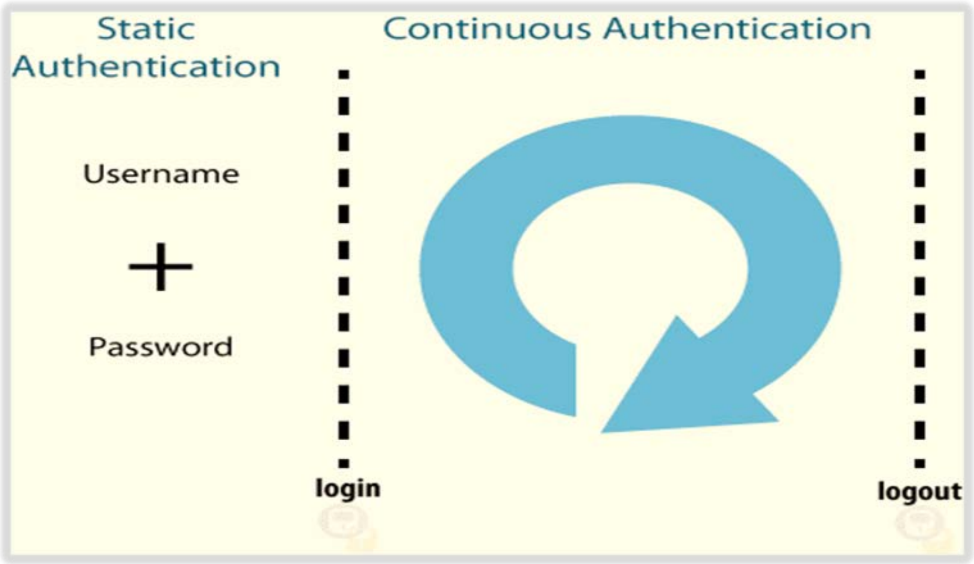


Figure #1: Static and continuous authentication [3]

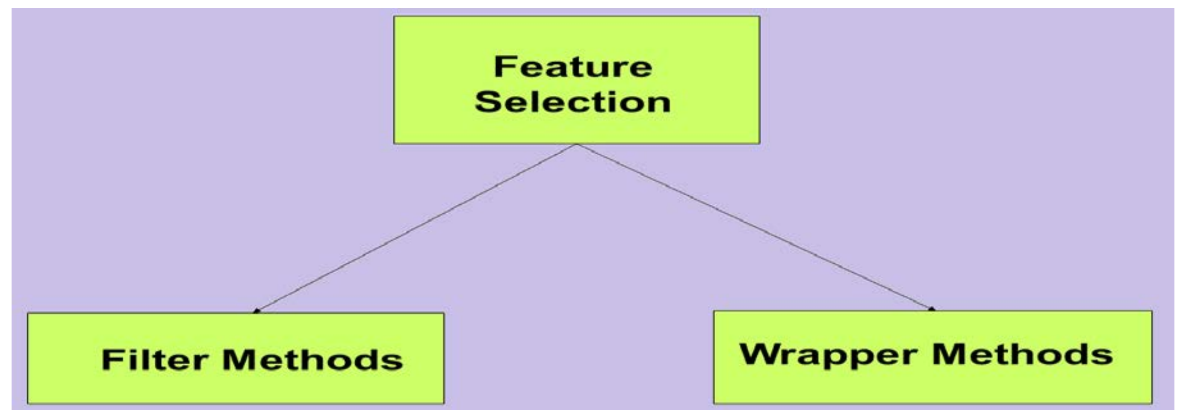
- Among all the biometric methods available, keystroke dynamics is selected because it is more reliable, less costly and faster than others techniques.
- Many machine learning algorithms have been used for keystroke dynamics analysis [1].
- It has been used in online examinations where the user is continuously verified during the exam. It has also been used in online banking for secure continuous verification of the user.

## Research Objective

- We strive to improve the accuracy of machine learning algorithms in keystroke dynamics analysis by using feature selection prior to using them on the keystroke dataset.

## Feature Selection

- The process of removing features from the data set that are irrelevant with respect to the task that is performed.[2]
- It reduces overfitting, improves accuracy and reduces training time.



Feature Selection

Filter Methods

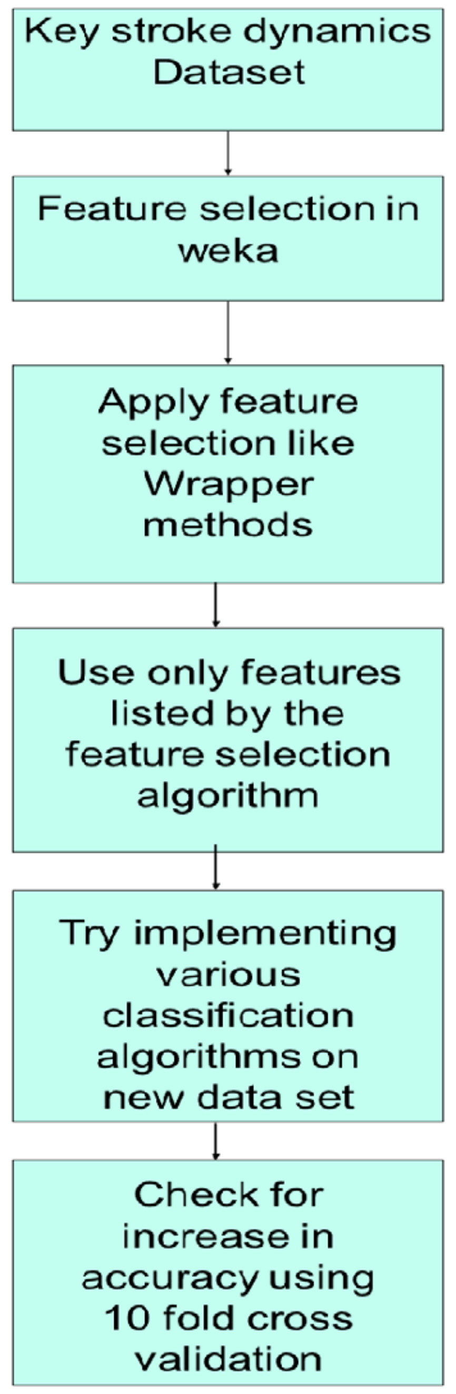
Wrapper Methods

Figure #2: Forms of feature selection

## Approach

The Weka tool is used with attribute selection. The process is separated into two parts:[2]

- Attribute Evaluator: Method by which attribute subsets are assessed. These include cfsSubsetEval, ClassifierSubsetEval, WrapperSubsetEval.
- Search Method: Method by which the space of possible subsets is searched. These include exhaustive, best first and greedy stepwise.



Key stroke dynamics Dataset

Feature selection in weka

Apply feature selection like Wrapper methods

Use only features listed by the feature selection algorithm

Try implementing various classification algorithms on new data set

Check for increase in accuracy using 10 fold cross validation

Figure #3 : Methodology

## Experimental Setup

- The experimental data [1] used was collected from 51 users each typing same password .tie5Roanl for 400 times. The various 31 timing features (e.g., the key down - key downtimes and hold times) were extracted.
- The data is used without any filtering or modification.
- The pre-process step of feature selection is conducted before using machine learning algorithm on the data set.
- During classification of the user based on the keystroke pattern, the major challenge was to split the dataset for classification.
- We have considered multi-class classification of 51 classes because it is easier to compare one subject to the other and try to authenticate each with respect to the other.

## Experimental Results

The following table shows the increase in accuracy for three machine learning algorithms after using feature selection with them.

| Machine Learning Algorithm | Feature Selection Method | Accuracy Before Feature Selection | Accuracy After Feature Selection |
|---|---|---|---|
| J48 | Wrapper Method-J48 algorithm | 82% | 83% |
| Random tree | Wrapper Method-Random Tree | 69.76% | 69.848% |
| SMO | Symmetrical Uncertain Attribute Eval | 80.039% | 82.243% |
| Decision Table | Wrapper method-J48 | 42.09% | 42.3% |
| IBK | Relief Attribute Eval | 78.5% | 80% |
| Hoeffding Tree | Relief Attribute Eval | 65.61% | 68.34% |

Figure #4: Results showing increase of accuracy after feature selection

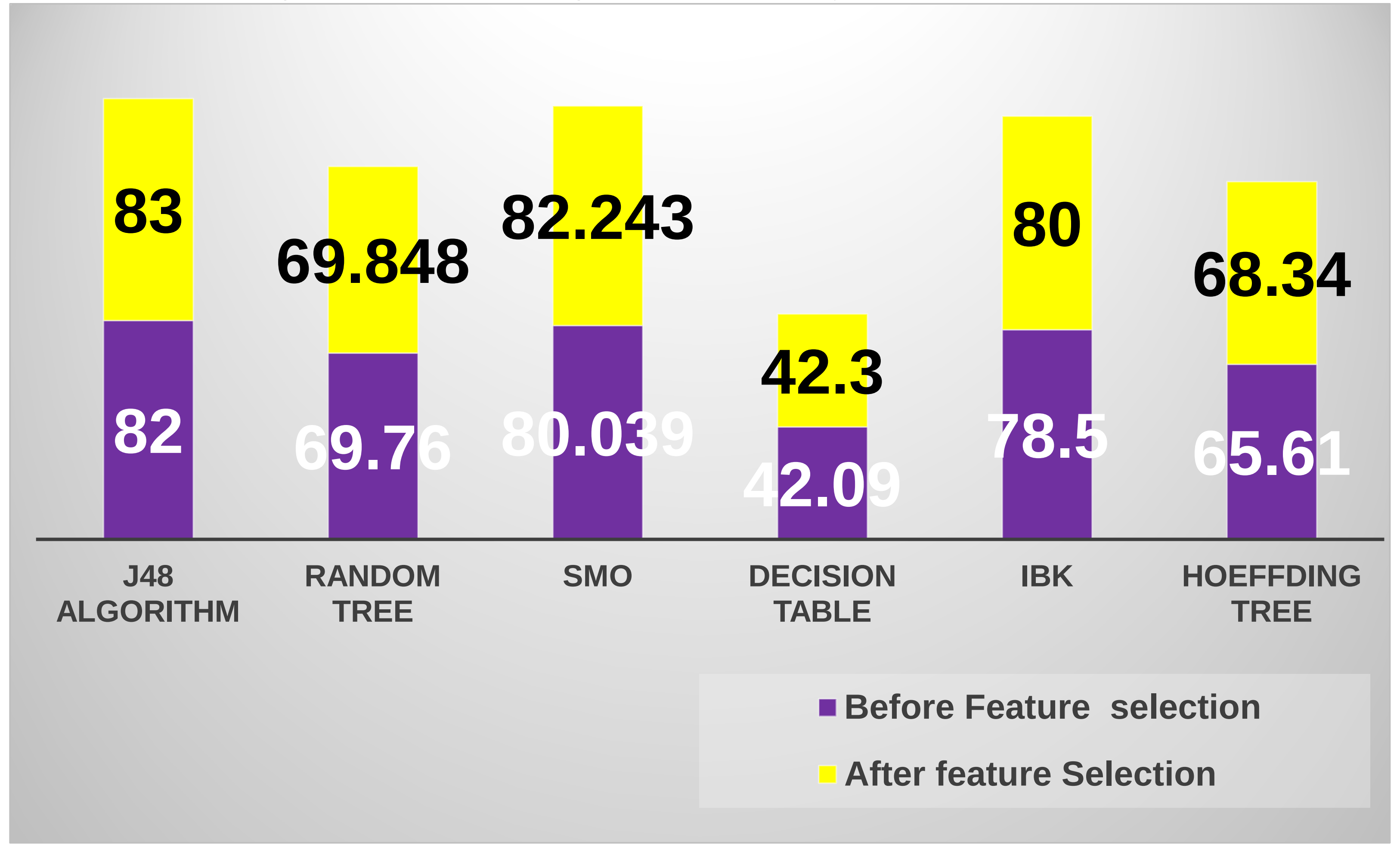


Figure #5 : Machine learning algorithm performance before and after feature selection

## Conclusion

Our experiments show that using feature selection along with machine learning, it is possible to increase the accuracy of authentication system to some extent. We continue to analyze various other combinations of algorithms with feature selection for better results.

## References

1. Kevin S. Killourhy and Roy A. Maxion. "Comparing Anomaly Detectors for Keystroke Dynamics," in Proceedings of the 39th Annual International Conference on Dependable Systems and Networks (DSN-2009), pages 125-134, Estoril, Lisbon, Portugal, June 29-July 2, 2009. IEEE Computer Society Press, Los Alamitos, California, 2009.
2. Doraisamy, Shyamala, et al. "A Study on Feature Selection and Classification Techniques for Automatic Genre Classification of Traditional Malay Music." *ISMIR*. 2008.
3. Victoria, B C. 2011. *What Is Continuous Authentication?*, PlurilockOn . https://goo.gl/xPDgs0