The deriving question for my research was “Is it possible to quantify the behavioural change of items in temporal data? And, what is the best reference point to compare the behaviour change with?”. So basically, this thesis can be considered as study in the data analysis field as it proposes two methods in this area to answer the primary question. One method is for classifying items’ behaviour in temporal data sets using temporal rule-based classification and another method to measuring the amount of changes in their behaviour.

To measure changes over time we proposed method which consists of two steps. In the first step, the temporal data will be converted into multiple consequent time points with each time point containing a single reading of for each item these items will be clustered using any traditional none-temporal clustering algorithm like k-means. Then in the next step changes of the items behaviour in these time points will be measured against a reference of behaviour using external cluster validity indices. The reference of behaviour can be the first time-point, previous time point or the overall general behaviour of the item in the temporal data. While multiple methods exist for measuring changes in data streams however these methods focus on the clusters and their movement instead of the items inside clusters.

The last reference of behaviour required classifying items in the temporal data sets.

The Temporal rule-based classification method consists of two steps Rule generation step and Rule optimisation. Classification rules for this method can be provided by filed experts with the help of profiling items in the data to see possible limits different classes. The rules are generated using aggregative measures of the temporal data, like their standard deviation, mean and/or sum of differences. Instead of providing crisp splitting values between classes, the rules can be provided in the form of flexible ranges of values between classes; this might result in a large number of crisp rules.

The second stage of the classification, optimises the pool of the rules by selecting the optimum classifier among all provided ranges of classifiers to find the most possible compacted class of items in each time point. To measure compactness, distance measures like Euclidean distance or statistical measures like standard deviation and variance are used to measure the distance among items of the same class in each time point. The optimisation goal is to minimise the overall distance among items of the same class in all time points. Here we used two methods for optimisation brute force and differential evolution.

The benefit of this method is that it provides simple classification rules that are easy to be understood and manage by experts of the field on the one hand, and taking advantage of the temporal dimension of the data to provide the optimum class for each item on the other hand.

To test the proposed method four different data sets are used

The first data set was a synthetic data set which successfully validated the measuring changes over time method.

Then two different data sets of the public goods game are used to classify players according to their behaviour during game rounds and to measure changes in their behaviour.

The results are compared with the available labels of players which are provided by the economists and other available classification methods like SVM. In both cases the proposed method proved to be a better alternative for the available methods of classification.

And we have compared the proposed method of measuring change over time with MONIC method. Both methods (the proposed and monic) confirmed the findings of the economists that the players behaviour are gradually changing over game rounds which might reflect the learning curve of the players.

Finally, we used a data set of S&P500 stock market to test the ability to generalise and scalability of the proposed methods. Stocks are classified into four classes according to their market price stability. The proposed algorithms proved to be scalable and flexible with different data sets. With our proposed methods we might provide a way for classifying stock market stability and we present another evidence for random walk case in the debate of stock market predictability.