

4<sup>th</sup> International Conference on

# BIG DATA ANALYSIS AND DATA MINING

September 07-08, 2017 | Paris, France

## Implementation and performance comparison of Wilson, holdout, and Multiedit algorithm for edited k-nearest neighbor

**Febriana Misdianti**

University of Manchester School of Computer Science, UK

**K**-nearest neighbor (k-NN) classifier is widely used for classifying data in various segments. However, k-NN classifier has high computational cost because it uses linear search through all training data. In naïve implementation, k-NN will go through all training set  $i$  to compute its distance  $d$  from input data ( $O(nd)$ ). Then, it loops again for all training set  $n$  to find  $k$  smallest results ( $O(nk)$ ). So, the overall time complexity for k-NN is  $O(nd+nk)$ . Thus, it is not suitable to classify multidimensional data with a huge number of training set. Meanwhile, k-NN needs a large number of samples in order to work well. There have been several ideas proposed to increase the time performance of k-NN in predicting a test data. One of the popular ideas is by reducing the number of TS in a model so that it would cut the testing time as the number of data that need to be explored becomes smaller. The aim of this experiment is to implement k-NN editing algorithms that cut the number of training data so that it become faster in predicting an input. This experiment implements three editing algorithms, namely, Wilson's editing, holdout editing, and Multiedit algorithm, and also compare the performance of them.

### Biography

Febriana Misdianti is a Post-graduate student at University of Manchester. She has completed her Bachelor degree of Computer Science at Universitas Indonesia and has two years of working experience in startup companies in Jakarta and Singapore. She has been winning a lot of competitions related to computer science and has published a paper about data security in a reputable journal.

febriana.misdianti@gmail.com

Notes: