



ABSTRACT

Naïve Bayes Classifier is a classification algorithm in data mining. Within the classification process, Bayesian theorem is adopted by Naïve Bayes Classifier to map a data into its class by measuring probability from its attributes. Currently, Naïve Bayes Classifier algorithm is based on probability distribution. However, the accuracy of Naïve Bayes Classifier is still able to be improved by measuring the weight of each attribute in the class.

This accuracy improvement is performed by measuring weight from each attribute, which is obtained from the value of R Square. By using this method, a new parameter is added for Naïve Bayes Classifier in calculating probability to reach the posterior probability, in which the parameter is R Square that is obtained from the relationship level between attribute and its class. Joint probability is used by conditional probability calculation method to prevent an error in commutative multiplication. Also, to prevent the occurrence of zero-probability, the proposed method is enhanced by Laplacian method.

This research produces formula of the new method from Naïve Bayes Classifier that is based on probability attribute and weighting attribute with R Square, which is called Correlated Naïve Bayes Classifier. Improvement on the accuracy is shown on Correlated Naïve Bayes Classifier algorithm compared to the previous algorithm, Naïve Bayes Classifier, using the tested dataset with the significant value of 14.03%. A comparison with the other methods that was also performed from the previous researches also shows that proposed method is more effective and efficient with the accuracy of 66.36% and complexity $O(n^2)$.

Keywords – data mining, classification, Naïve Bayes Classifier, Correlated Naive Bayes Classifier, weighting, laplacian.



INTISARI

Naïve Bayes Classifier merupakan salah satu algoritme klasifikasi dalam *data mining*. Dalam proses klasifikasi, *Naïve Bayes Classifier* mengadopsi teorema *Bayesian* untuk memetakan suatu data terhadap kelas dengan memperhitungkan *probability* dari atribut data tersebut. Sampai saat ini, algoritma *Naïve Bayes Classifier* hanya berdasar pada distribusi probabilitas. Salah satu hal yang berpotensi untuk meningkatkan akurasi dari *Naïve Bayes Classifier* adalah dengan memperhitungkan bobot dari masing-masing atribut terhadap kelas.

Peningkatan akurasi *Naïve Bayes Classifier* dilakukan dengan memperhitungkan bobot dari masing-masing atribut yang didapatkan dari nilai *R Square* dari masing-masing atribut terhadap kelas. Dengan metode ini, *Naïve Bayes Classifier* menambahkan satu parameter tambahan dalam perhitungan *probability* untuk mencapai *posterior probability* yaitu *R Square* yang didapat dari tingkat hubungan antara atribut dengan kelas. Metode perhitungan *conditional probability* digunakan *joint probability* untuk mencegah terjadinya kesalahan komutatif perkalian. Kemudian untuk mencegah terjadinya *probability 0*, metode yang diusulkan disempurnakan dengan menggunakan metode *Laplacian*.

Hasil dari penelitian ini adalah sebuah rumusan metode baru dari algoritma *Naïve Bayes Classifier* yang berbasis pada *probability attribute* dan bobot atribut dengan *R Square* yang dinamakan *Correlated Naïve Bayes Classifier*. Algoritme *Correlated Naïve Bayes Classifier* memberikan kenaikan akurasi dibandingkan algoritme *Naïve Bayes Classifier* dengan data set yang diuji sebesar 14,03% secara *significant*. Hasil perbandingan dengan metode-metode pembobotan lainnya yang telah dilakukan oleh peneliti sebelumnya juga menunjukkan bahwa metode yang diusulkan lebih efektif dan efisien dengan tingkat akurasi 66,36 % dan kompleksitas proses $O(n^2)$.

Kata kunci – *Data mining*, klasifikasi, *Naïve Bayes Classifier*, *Correlated Naïve Bayes Classifier*, pembobotan, *Laplacian*.