

CHAPTER 6 CONCLUSION

Based on the results of research on Comparing the KNN and Random Forest Algorithms in the classification of student data, it can be concluded that:

1. The main steps for the classification of the Random Forest Algorithm and the KNN are dividing the data based on the probability that training data and testing data appear. The classification of the KNN algorithm is based on the distance to the nearest neighbor. While the classification of the Random Forest Algorithm is based on the voting results of the formed tree.

2. The comparison between the KNN algorithm and the Random Forest algorithm based on the precision value without adding a dataset is that the precision value of the KNN algorithm is higher than the random forest algorithm in split data testing by 60%, 70%, 80%, and 90%. On the other hand, the random forest algorithm gets the highest value on the split data test by 50%. The smaller the False Positive value, the greater the precision value.

3. The comparison based on the recall value of the two algorithms without adding a new dataset is that the KNN algorithm gets a higher value than the random forest algorithm with split data testing of 50%, 60%, 70%, 80%, 90%. The smaller the False Negative value, the greater the recall value.

4. Comparison based on the f1 score or the average of the precision and recall of the two algorithms is that the KNN algorithm is better in testing split data by 50%, 60%, 70%, 80%, and 90%.

5. The addition of a new dataset proves that the value of precision, recall and f1 score in both algorithms has increased due to the amount of data being trained is the same and the values of tp, tn, fp, fn are the same.

Based on the above conclusions, the following suggestions are given for future research:

1. It is recommended for further research to use and apply other algorithms such as the Kmeans algorithm, the Naive Bayes algorithm and so on in getting even better accuracy values.

2. It is expected to use more complex variables such as student graduation rates based on student batches.

3. It is hoped that future research can be developed and become a reference in developing data classification.

4. Using real data so as to avoid the risk of data similarity in calculating accuracy.

