

## **BAB V**

### **PENUTUP**

#### **5.1 Kesimpulan**

Berdasarkan hasil analisis data dan pembahasan diperoleh kesimpulan sebagai berikut:

1. Dari 15 atribut pada penelitian ini, terdapat 4 atribut yang dianggap berpengaruh untuk klasifikasi lama studi mahasiswa. Atribut berpengaruh berturut-turut yakni, Waktu studi dengan nilai pengaruh sebesar (0,990), Asal sekolah (0,763), Tempat lahir (0,736), dan Dosen Penasehat Akademik sebesar (0,510).
2. Pengklasifikasian lama Studi Mahasiswa Pendidikan Matematika angkatan 2013 dengan menggunakan algoritma *K-NN* diperoleh hasil yakni dari 10 orang dengan kelas tepat waktu dan tidak terdapat kesalahan dalam klasifikasi. Sedangkan dari 5 orang yang lewat batas waktu, terdapat 3 orang yang dapat diklasifikasikan dengan benar dan 2 orang lainnya di klasifikasikan tepat waktu dengan demikian terdapat 2 orang yang tidak dapat diklasifikasikan dengan benar. Selanjtnya tingkat akurasi dari algoritma *K-NN* sebesar 86,67%.
3. Pengklasifikasian lama Studi Mahasiswa Pendidikan Matematika angkatan 2013 dengan menggunakan algoritma *random forest* diperoleh hasil yakni dari 10 orang dengan kelas tepat waktu dan tidak terdapat kesalahan dalam klasifikasi. Dan juga dari 5 orang yang lewat batas waktu, tidak terdapat kesalahan. Sehingga dapat dilihat tingkat akurasi dengan algoritma *random forest* sebesar 100%.

## 5.2 Saran

Saran yang dapat diberikan pada penelitian ini berdasarkan analisis dan kesimpulan sebagai berikut :

1. Pada penelitian ini menggunakan seleksi fitur *information gain*, diharapkan pada penelitian selanjutnya dapat menggunakan seleksi fitur yang lain.
2. Penerapan algoritma klasifikasi lainnya, seperti: SVM, C 4.5, Nave Bayes, ANN, dsb yang selanjutnya dapat dikomparasi untuk klasifikasi lama studi mahasiswa.
3. Dalam penelitian ini difokuskan pada klasifikasi, pada penelitian selanjutnya diharapkan dapat melakukan prediksi lama studi mahasiswa.

## DAFTAR PUSTAKA

- Adnyana, I. M. B. (2016). Prediksi Lama Studi Mahasiswa Dengan Metode Random Forest (Studi Kasus: STIKOM Bali). *CSRID (Computer Science Research and Its Development Journal)*, 8(3), 201-208.
- Agusta, L. (2009). Perbandingan algoritma stemming Porter dengan algoritma Nazief & Adriani untuk stemming dokumen teks bahasa indonesia. *Konferensi Nasional Sistem dan Informatika*, 2009, 196-201.
- Asha Gowda Karegowda, M.A. Jayaram, & A.S. Manjunath. (2012). Cascading k-means clustering and k-nearest neighbor classifier for categorization of diabetic patients. *International Journal of Engineering and Advanced Technology*, 1(3), 147151.
- Badu, Z. S. (2016). Penerapan Algoritma K-Nearest Neighbor Untuk Klasifikasi Dana Desa. *Jurnal Informatika*, November.
- Biau, Gerard. 2012. Analysis of a random forests model, *Journal of Machine Learning Research*, Vol. 13, pp. 1063-1095
- Bimantoro, D. A., & Uyun, S. . (2017). Pengaruh Penggunaan Information Gain untuk Seleksi Fitur Citra Tanah Dalam Rangka Menilai Kesesuaian Lahan Pada Tanaman Cengkeh. *Jiska*, 2(1), 4252.
- Breiman, L., & Cutler, A. (2011). *Manual-Setting Up, Using, and Understanding Random Forests V4. 0*. 2003.
- Breiman, L. (1999). *Randon Forests*. *Machinelearning202.Pbworks.Com*, 135.

- Budiman, A. S., & Parandani, X. A. (2018). Uji akurasi klasifikasi dan validasi data pada penggunaan metode membership function dan algoritma c4. 5 dalam penilaian penerima beasiswa. *Simetris: Jurnal Teknik Mesin, Elektro dan Ilmu Komputer*, 9(1), 565-578.
- Budi, S. (2007) .Data Mining Teknik Pemanfaatan Data untuk Keperluan Bisnis. Teori dan Aplikasi . Yogyakarta, Graha Ilmu.
- Chormunge, S., & Jena, S. (2016). Efficient Feature Subset Selection Algorithm for High Dimensional Data. *International Journal of Electrical & Computer Engineering* (2088-8708), 6(4).
- Crone, S. F., Lessmann, S., & Stahlbock, R. (2006). The impact of preprocessing on data mining: An evaluation of classifier sensitivity in direct marketing. *European Journal of Operational Research*, 173(3), 781-800.
- Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge university press.
- Francisco, A. R. L. (2013). The Top Ten Algorithms in Data Mining, Chapman & Hall/CRC Data Mining and Knowledge Discovery. In *Journal of Chemical Information and Modeling* (Vol. 53, Issue 9).
- Gorunescu, F. (2011). *Data Mining: Concepts, models and techniques* (Vol. 12). Springer Science & Business Media.
- Han, J., Pei, J., & Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Han, J., Kamber, M., & Pei, J. (2012). *Data Mining: Concepts and Techniques* . USA: Morgan Kaufmann.
- Hardjianto, M., & Winarko, E. 2013. Penentuan Kesehatan Lansia Berdasarkan Multi Variabel Dengan Algoritma K-NN Pada Rumah Cerdas. *Prosiding Seminar Nasional Informatika*, Vol. 7, Hal. 214-218.

- Hartono, S. (2020). Journal Of Technology Information. Journal Of Technology Information, 5(1), 2530.
- Hassanat, A. B., Abbadi, M. A. dan Altarawneh, G. A., 2014. Solving the problem of the K parameter in the K-NN Classifier using an Ensemble learning Approach. Computer Science and Informatic, 12(8), pp. 33-39.
- Indonesia, T. P. K. B. B. (2008). Kamus besar bahasa Indonesia. Jakarta: Balai Pustaka.
- Jin, Z., Shang, J., Zhu, Q., Ling, C., Xie, W., & Qiang, B. (2020, October). RFRSF: Employee Turnover Prediction Based on Random Forests and Survival Analysis. In International Conference on Web Information Systems Engineering (pp. 503-515). Springer, Cham.
- KBBI. 2020. Kamus Besar Bahasa Indonesia (KBBI). [Online] Available at: <https://kbbi.kemdikbud.go.id> [Diakses 20 Juni 2020]
- Larose, D. T., & Larose, C. D. (2014). Discovering knowledge in data: an introduction to data mining (Vol. 4). John Wiley & Sons.
- Lestari, M. E. I. (2014). Penerapan algoritma klasifikasi Nearest Neighbor (K-NN) untuk mendeteksi penyakit jantung. Faktor Exacta, 7(4), 366-371.
- Maulana, M. R., & Al Karomi, M. A. (2015). Information Gain untuk Mengetahui Pengaruh Atribut Terhadap Klasifikasi Persetujuan Kredit. JURNAL PEMERINTAH KOTA PEKALONGAN, 9.
- PERMENDIKNAS. (2000). Keputusan Menteri Pendidikan Nasional Republik Indonesia No.232. Jakarta: Depdiknas
- Prasetyo, E. (2012). Data mining konsep dan aplikasi menggunakan matlab. Yogyakarta: Andi.

- Ratnawati, L., & Sulistyningrum, D. R. (2020). Penerapan Random Forest untuk Mengukur Tingkat Keparahan Penyakit pada Daun Apel. *Jurnal Sains dan Seni ITS*, 8(2), A71-A77.
- Roihan, A. (2018). Seleksi fitur menggunakan Symmetrical Uncertainty pada prediksi cacat perangkat lunak (Doctoral dissertation, Universitas Islam Negeri Maulana Malik Ibrahim).
- Rokach, L., & Maimon, O. (2010). *Data Mining and Knowledge Discovery Handbook* Second edition pages 167-192.
- Shaltout, N. A., El-Hefnawi, M., Rafea, A., & Moustafa, A. (2014). Information gain as a feature selection method for the efficient classification of influenza based on viral hosts. In *Proceedings of the world congress on engineering* (Vol. 1, No. 2014, pp. 625-631).
- Subrata, K. K. A., Widyantara, I. M. O., & Linawati, L. (2016). Klasifikasi Penggunaan Protokol Komunikasi Pada Trafik Jaringan Menggunakan Algoritma K-Nearest Neighbor. *Majalah Ilmiah Teknologi Elektro*, 16(1), 67.
- Untari, D. (2014). *Data Mining Untuk Menganalisa Prediksi Mahasiswa Berpotensi Non-Aktif Menggunakan Metode Decision Tree C4. 5*. Semarang. Universitas Dian Nuswantoro.
- Vijayakumar, V., & Nedunchezian, R. (2012). A study on video data mining. *International journal of multimedia information retrieval*, 1(3), 153-172.
- Witten, I. F. E., & Hall.(2011). *Data Mining: Practical Machine Learning and tools*.
- Wu, X., & Kumar, V. (Eds.). (2009). *The top ten algorithms in data mining*. CRC press.

