BAB 6

KESIMPULAN DAN SARAN

6.1 Kesimpulan

Kesimpulan yang dapat diambil adalah sebagai berikut

1. Berdasarkan hasil pengujian dan perhitungan data, akurasi yang didapat dengan menggunakan metode ini bisa mencapai 99%.

2. Untuk dapat menemukan anomali, model *machine learning* yang digunakan oleh penulis adalah sebagai berikut:

   (a) Metrik perhitungan jarak (*Distance Similarity*) yang digunakan adalah *Euclidean Distance*

   (b) Klasifikasi menggunakan K-NN dengan nilai K = 3

   (c) Bantuan *clusering* menggunakan DBSCAN memiliki attribut minimal titik adalah 3, jarak minimal adalah 1.3

3. Pembangunan aplikasi pendeteksi anomali menggunakan *machine learning* sudah berhasil dibangun. Aplikasi dibangun menggunakan bantuan dari JavaFX. Aplikasi dapat memilih *file* yang digunakan sebagai data untuk diberi label ataupun untuk diprediksi. Untuk dapat memprediksi menggunakan aplikasi ini, sebelumnya sudah memiliki data yang diberli label (data *train*) dan sudah tersimpan didalam aplikasi. Penyimpanan data *train* menggunakan *file* berformat .txt. Pada aplikasi diimplementasikan Z-Score, K-NN dan DBSCAN.

4. Berdasarkan hasil pengujian, Klasifikasi menggunakan K-NN dapat digunakan untuk memprediksi log yang bersifat anomali dengan akurasi 99,42%. Hal ini dapat terjadi dikarenakan klasifikasi memiliki data yang sudah dilabeli terlebih dahulu. Namun prediksi yang dilakukan dikhususkan untuk log yang berasal dari sumber yang sama dengan log *training*. Hal ini disebabkan karena setiap *website* memiliki fungsi dan karakteristik yang berbeda - beda dari setiap *request HTTP* yang diberikan, sehingga kurang cocok jika menyilangkan data *train* dan prediksi dari berbagai website yang ada.

6.2 Saran

Saran dari penulis untuk penelitian atau penambahan selanjutnya adalah sebagai berikut:

1. Dikarenakan fitur yang diolah masih banyak sehingga proses perhitungan menjadi kompleks, maka disarankan mencari atau mengubah fitur yang sudah digunakan agar dapat memperingan proses perhitungan dari *machine learning* ataupun melakukan reduksi dimensi menggunakan teknik -teknik tertentu.

2. Menggunakan algortima *machine learning* atau fitur yang lain agar log - log yang digunakan tidak terikat dari sifat website, namun bisa dikelola secara universal.

3. Mengembangkan perangkat lunak yang dibuat agar dapat mendeteksi log selain dari apache, namun bisa dari webserver yang lain seperti *nginx*. Juga mengembangkan perangkat lunak agar lebih banyak fungsionalitasnya agar pengguna dapat lebih nyaman menggunakan perangkat lunak ini.

DAFTAR REFERENSI

[1] Apache access log description. https://httpd.apache.org/docs/2.4/logs.html. Accessed: 2019-09-10.

[2] Wang, S., Bi, J., Wu, J., Yang, X., dan Fan, L. (2012) On adapting HTTP protocol to content centric networking. *Proceedings of the 7th International Conference on Future Internet Technologies*, New York, NY, USA CFI '12, pp. 1–6. ACM.

[3] Chandola, V., Banerjee, A., dan Kumar, V. (2009) Anomaly detection: A survey. *ACM computing surveys (CSUR)*, **41**, 15.

[4] Halfond, W. G., Viegas, J., Orso, A., dkk. (2006) A classification of sql-injection attacks and countermeasures. *Proceedings of the IEEE international symposium on secure software engineering*, pp. 13–15. IEEE.

[5] Spett, K. (2005) Cross-site scripting. *SPI Labs*, **1**, 1–20.

[6] Garcia, V. H., Monroy, R., dan Quintana, M. (2006) Web attack detection using id3. *IFIP World Computer Congress, TC 12*, pp. 323–332. Springer.

[7] Alpaydin, E. (2009) *Introduction to machine learning*. MIT press.

[8] Kotsiantis, S. B., Zaharakis, I., dan Pintelas, P. (2007) Supervised machine learning: A review of classification techniques. *Emerging artificial intelligence applications in computer engineering*, **160**, 3–24.

[9] Aggarwal, C. C. (2014) *Data classification: algorithms and applications*. CRC press.

[10] Bhargava, N., Sharma, G., Bhargava, R., dan Mathuria, M. (2013) Decision tree analysis on j48 algorithm for data mining. *Proceedings of International Journal of Advanced Research in Computer Science and Software Engineering*, **3**.

[11] Wettschereck, D., Aha, D. W., dan Mohri, T. (1997) A review and empirical evaluation of feature weighting methods for a class of lazy learning algorithms. *Artificial Intelligence Review*, **11**, 273–314.

[12] Daumé III, H. (2012) A course in machine learning. *Publisher, ciml. info*, **5**, 69.

[13] Gan, G., Ma, C., dan Wu, J. (2007) *Data clustering: theory, algorithms, and applications*. Siam.

[14] Hartigan, J. A. dan Wong, M. A. (1979) Algorithm as 136: A k-means clustering algorithm. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, **28**, 100–108.

[15] Alasadi, S. A. dan Bhaya, W. S. (2017) Review of data preprocessing techniques in data mining. *Journal of Engineering and Applied Sciences*, **12**, 4102–4107.

[16] Kandel, S., Heer, J., Plaisant, C., Kennedy, J., Van Ham, F., Riche, N. H., Weaver, C., Lee, B., Brodbeck, D., dan Buono, P. (2011) Research directions in data wrangling: Visualizations and transformations for usable and credible data. *Information Visualization*, **10**, 271–288.

[17] Raschka, S. (2018) Model evaluation, model selection, and algorithm selection in machine learning. *arXiv preprint arXiv:1811.12808* , **?**

[18] Baratloo, A., Hosseini, M., Negida, A., dan El Ashal, G. (2015) Part 1: simple definition and calculation of accuracy, sensitivity and specificity. , **?**