

## BAB 6

### KESIMPULAN DAN SARAN

Bab ini berisi kesimpulan dari awal hingga akhir penelitian beserta saran untuk penelitian selanjutnya.

#### 6.1 Kesimpulan

Kesimpulan yang dapat ditarik dari penelitian ini adalah sebagai berikut:

- Pada penelitian ini, telah dipelajari dua algoritma *k-anonymity* yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Dalam penelitian ini juga mempelajari bagaimana cara kerja algoritma *Datafly* dan algoritma *Greedy k-member Clustering* dalam menganonimisasi data.
- Pada penelitian ini, telah diimplementasikan dua algoritma *k-anonymity* yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Dalam penelitian ini, telah dibangun perangkat lunak yang mengimplementasikan algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Selain menghasilkan anonimisasi data, nilai  $k$  juga dapat diatur pada perangkat lunak ini, sesuai dengan kebutuhan pengguna. Perangkat lunak juga dapat menyimpan hasil anonimisasi data ke dalam *format .zip*.
- Pada penelitian ini, telah dilakukan pengujian fungsional pada perangkat lunak untuk mendapatkan hasil yang sesuai dengan yang diinginkan yang sesuai dengan tujuan dari *k-anonymity*.
- Pada penelitian ini, dilakukan juga pengujian eksperimental untuk mendapatkan relasi antara *information loss*, nilai  $k$ , tipe dan banyaknya atribut yang digunakan, dan waktu eksekusi algoritma *k-anonymity* yang diimplementasikan. Dari pengujian ini, disimpulkan bahwa setiap algoritma memiliki kelebihan dan kelemahannya masing-masing, seperti algoritma *Datafly* memiliki *information loss* dan waktu eksekusi yang rendah bila nilai  $k$  semakin tinggi, tetapi jumlah *tuple* pada hasil anonimisasi menjadi lebih sedikit dibanding dengan jumlah *tuple* pada tabel privat. Tetapi, algoritma *Greedy k-member Clustering* memiliki *information loss* dan waktu eksekusi yang tinggi bila nilai  $k$  semakin tinggi, tetapi *tuple* pada hasil anonimisasi sama dengan jumlah *tuple* pada tabel privat.
- Pada penelitian ini, pada saat pengujian eksperimental dapat disimpulkan juga bahwa *information loss* dapat diatur dengan menyesuaikan nilai  $k$  dengan algoritma yang digunakan atau menyesuaikan jumlah atribut yang digunakan. Nilai  $k$  dan banyaknya atribut yang digunakan juga dapat berpengaruh pada waktu eksekusi algoritma *k-anonymity* yang diimplementasikan.
- Pada penelitian ini, pada saat pengujian eksperimental nilai  $k$  mempengaruhi jumlah *tuples* pada hasil anonimisasi yang *cluster*-nya sama dengan *tuples* pada tabel privat, tetapi setiap kasus memiliki nilai  $k$  yang berbeda agar jumlah *tuples* pada hasil anonimisasi yang *cluster*-nya sama dengan *tuples* pada tabel privat maksimal.

## 6.2 Saran

Saran untuk penelitian selanjutnya adalah sebagai berikut:

- Pada penelitian ini, perangkat lunak yang dibuat hanya sebatas mengimplementasikan dua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Pada algoritma tersebut, masih terdapat bagian-bagian yang belum optimal. Untuk penelitian selanjutnya, algoritma *k-anonymity* yang diimplementasikan dapat dioptimalkan kembali agar waktu eksekusi algoritma *k-anonymity* yang diimplementasikan lebih cepat dari yang sebelumnya.
- Pada penelitian ini, perangkat lunak yang dibuat hanya mengimplementasikan dua algoritma *k-anonymity*, yaitu algoritma *Datafly* dan algoritma *Greedy k-member Clustering*. Karena algoritma *k-anonymity* merupakan salah satu algoritma anonimisasi yang populer, terdapat beberapa algoritma lain yang bisa memenuhi tujuan dari *k-anonymity*. Oleh karena itu, untuk penelitian selanjutnya dapat mengimplementasikan algoritma lain yang memenuhi tujuan *k-anonymity* ke dalam perangkat lunak yang sudah dibuat.
- Pada penelitian ini, studi yang hanya berfokus pada tujuan *k-anonymity* saja. Oleh karena itu, penelitian selanjutnya dapat mempelajari ekspansi dari *k-anonymity*, seperti *t-closeness* dan *l-diversity* agar hasil anonimisasi data semakin baik untuk dilakukan *data mining*.
- Pada penelitian ini, pengujian yang dilakukan masih sangat sedikit karena hanya menggunakan 1000 buah *record* saja. Untuk penelitian selanjutnya, pengujian yang dilakukan bisa menggunakan data yang lebih besar lagi dibandingkan dengan pengujian pada penelitian ini.

## DAFTAR REFERENSI

- [1] Han, J., Pei, J., dan Kamber, M. (2011) *Data mining: concepts and techniques*. Elsevier.
- [2] Moore, A. D. (2003) Privacy: Its meaning and value. *American Philosophical Quarterly*, **40**, 215–227.
- [3] Mendes, R. dan Vilela, J. P. (2017) Privacy-preserving data mining: methods, metrics, and applications. *IEEE Access*, **5**, 10562–10582.
- [4] Sweeney, L. (2002) k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, 557–570.
- [5] McCallister, E. (2010) *Guide to protecting the confidentiality of personally identifiable information*. Diane Publishing.
- [6] Berkhin, P. (2006) A survey of clustering data mining techniques. *Grouping multidimensional data*, pp. 25–71. Springer.
- [7] Sweeney, L. (2002) Achieving k-anonymity privacy protection using generalization and suppression. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, **10**, 571–588.
- [8] Byun, J.-W., Kamra, A., Bertino, E., dan Li, N. (2007) Efficient k-anonymization using clustering techniques. *International Conference on Database Systems for Advanced Applications*, pp. 188–200. Springer.
- [9] Shafranovich, Y. (2005) Common format and mime type for comma-separated values (csv) files. RFC 4180. RFC Editor, <http://www.rfc-editor.org>.
- [10] (2005) *The Oxford Handbook of Computational Linguistics (Oxford Handbooks)*. Oxford University Press.