

Privacy Preserving Data Publishing untuk Multiple Sensitive Attributes dengan Overlapped Slicing = Privacy Preserving Data Publishing with Multiple Sensitive Attributes using Overlapped Slicing

Widodo, author

Deskripsi Lengkap: <https://lib.ui.ac.id/detail?id=20506034&lokasi=lokal>

Abstrak

Privacy preserving data publishing (PPDP) merupakan bidang yang saat ini berkembang dengan fokus penelitian adalah mempertahankan data agar bersifat privat jika data tersebut dipublikasikan. Isu penting pada penelitian PPDP adalah meminimalkan nilai information loss yang diperoleh akibat proses penganoniman tabel mikrodata sehingga menjadi lebih privat. Berbagai model dan metode telah dikembangkan untuk mengatasi permasalahan tersebut. Model seperti k-anonymity, l-diversity, dan p-sensitive menjadi model dasar atas berkembangnya disiplin ilmu ini. Namun sebagian besar penelitian lebih banyak berfokus pada model untuk single sensitive attribute atau satu atribut sensitif pada tabel mikrodata. Padahal dalam dunia nyata atribut sensitif pada sebuah tabel bisa banyak atau multiple sensitive attributes. Penelitian yang membahas multiple-sensitive attributes pun masih banyak permasalahan yang belum terpecahkan karena hanya bertujuan untuk mengatasi satu permasalahan tertentu saja, misalnya untuk mengatasi serangan tertentu terhadap data. Sementara itu efek information loss kurang diperhatikan. Hal lain yang belum terlalu diperhatikan adalah bagaimana mendistribusikan nilai atribut sensitif ke seluruh grup data. Pendistribusian ini sangat penting untuk menghindari penumpukan data sensitif pada sebuah atau beberapa grup saja. Penelitian ini berhasil mengusulkan dan mengevaluasi model PPDP dengan overlapped slicing pada multiple sensitive attributes dengan metode pendistribusian nilai atribut sensitif berupa simple distribution of sensitive values (SDSV) dan extended systematic clustering (ESC). Penelitian ini juga mengusulkan sebuah pengukuran untuk menyempurnakan model pengukuran sebelumnya yaitu normalized and average discernibility metrics (NADM). Hasil dari penelitian ini menunjukkan overlapped slicing dengan tiga variasi metode untuk mencapai model tersebut memiliki tingkat information loss yang minimal dibandingkan dengan yang lain. Overlapped slicing dengan menggunakan variasi systematic clustering, SDSV, dan extended systematic clustering berhasil menghasilkan PPDP dengan nilai information loss yang kecil. Demikian juga dibandingkan dengan model lain yang menggunakan multiple sensitive attributes, overlapped slicing memiliki nilai information loss yang lebih kecil. Pada saat dijalankan dengan adult dataset, Nilai information loss yang telah dinormalkan untuk overlapped slicing adalah 0.25, sedangkan systematic clustering 0.625, SDSV 0.871, dan ESC 0,704. Dengan data bank marketing overlapped slicing menghasilkan nilai information loss yang dinormalkan sebesar 0.397, lebih baik daripada systematic clustering 0.441.

Privacy preserving data publishing (PPDP) is a field with research focus is in maintaining data to be private when the data is published. An important issue in PPDP is minimizing the information loss that is obtained due to the anonymization process to the microdata table so that it becomes more private. Various models and methods have been developed to overcome these problems. Models such as k-anonymity, l-diversity, and p-sensitive are the basic models for the development of this discipline. However, most studies focus on models for single sensitive attributes in microdata table. Yet in the real world, sensitive attributes on a table can be multiple sensitive attributes. There are still many problems in research that discusses multiple-sensitive attributes, and it still has not been solved because it

only aims to overcome one particular problem for each research, for example to overcome certain attacks on data. Meanwhile the effect of information loss is less noticed. Another thing that has not been given much attention is how to distribute sensitive attribute values across data groups. This distribution is very important to avoid the accumulation of sensitive data on just one or a few groups. This study successfully proposes and evaluates the PPDP model with overlapped slicing on multiple sensitive attributes and proposes methods for distributing sensitive attribute values namely, simple distribution of sensitive values (SDSV) and extended systematic clustering (ESC). This study also proposes a measurement to perfect the previous measurement model, normalized and average discernibility metrics (NADM). The results of this research show that overlapped slicing with three variation methods in achieving the model, has a minimal information loss compared to the others. Overlapped slicing by using systematic clustering, SDSV, and extended systematic clustering succeeded in producing PPDP with a small value of information loss. Likewise, compared to other models that use multiple sensitive attributes, overlapped slicing has a smaller information loss. When it is tested with adult dataset, the value of information loss that has been normalized for overlapped slicing is 0.25, while systematic clustering is 0.625, SDSV 0.871, and ESC is 0.704. With marketing bank dataset, it produces a normalized information loss value of 0.397, better than systematic clustering 0.441.