1. Record Nr. UNINA9910595049203321

Titolo	Privacy in statistical databases : International Conference, PSD 2022, Paris, France, September 21-23, 2022, proceedings / / Josep Domingo-Ferrer and Maryline Laurent, editors
Pubbl/distr/stampa	Cham, Switzerland : , : Springer Nature Switzerland AG, , [2022] ©2022
ISBN	3-031-13945-3
Descrizione fisica	1 online resource (375 pages)
Collana	Lecture notes in computer science ; ; 13463
Disciplina	005.8
Soggetti	Data protection Database security
Lingua di pubblicazione	Inglese
Formato	Materiale a stampa
Livello bibliografico	Monografia
Nota di bibliografia	Includes bibliographical references and index.
Nota di contenuto	Intro -- Preface -- Organization -- Contents -- Privacy Models -- An Optimization-Based Decomposition Heuristic for the Microaggregation Problem -- 1 Introduction -- 2 The Decomposition Heuristic -- 3 The Local Search Improvement Heuristic -- 4 The Mixed Integer Linear Optimization Algorithm Based on Column Generation -- 5 Computational Results -- 6 Conclusions -- References -- Privacy Analysis with a Distributed Transition System and a Data-Wise Metric -- 1 Introduction -- 2 Distributed Labeled-Tagged Transition Systems -- 3 -Indistinguishability, -Local-Differential Privacy -- 4 -Differential Privacy -- 5 Comparing Two Nodes on One or More Runs -- 6 New Metric for Indistinguishability and DP -- 7 Related Work and Conclusion -- References -- Multivariate Mean Comparison Under Differential Privacy -- 1 Introduction -- 2 Mathematical Background -- 2.1 Statistical Tests for Two Samples -- 2.2 Hotelling's t2-Test -- 2.3 Differential Privacy -- 3 Privatized Mean Comparison -- 3.1 Privatization of the t2-Statistic -- 3.2 Bootstrap -- 4 Simulation -- 5 Conclusion -- A Proofs -- B Effects of Privatization - Example -- C Algorithms -- References -- Asking the Proper Question: Adjusting Queries to Statistical Procedures Under Differential Privacy -- 1 Introduction -- 1.1 Setting -- 1.2 Our Contribution -- 1.3 Related Work -- 2 Fixed (Non-random) Datasets -- 2.1 Confidence Regions --

2.2 Testing Hypotheses: Likelihood-Ratio Test -- 3 Random, Normally Distributed Data -- 3.1 Confidence Regions -- 3.2 Testing Hypotheses: Likelihood-Ratio Test -- 4 Numerical Example -- 5 Appendix -- References -- Towards Integrally Private Clustering: Overlapping Clusters for High Privacy Guarantees -- 1 Introduction -- 2 Preliminaries -- 2.1 Integral Privacy -- 2.2 k-Anonymity, Microaggregation, and MDAV -- 2.3 Genetic Algorithms -- 3 -Centroid c-Means.
3.1 Formalization -- 3.2 Properties -- 4 Experiments -- 4.1 Solving the Optimization Problem -- 4.2 Datasets -- 4.3 Parameters -- 4.4 Results -- 5 Conclusions -- References -- II Tabular Data -- Perspectives for Tabular Data Protection - How About Synthetic Data? -- 1 Introduction -- 2 Recalling the Methods under Consideration -- 2.1 Synthetic Data -- 2.2 Targeted Record Swapping (TRS) -- 2.3 CKM Noise Design for Tabulations of Continuous Variables -- 3 Study Design -- 3.1 Test Data -- 3.2 Application Settings for Synthetic Data Generation -- 3.3 Application Settings for Targeted Record Swapping -- 3.4 Application Settings for the Cell Key Method -- 4 Measuring Utility and Disclosure Risk -- 5 Results -- 5.1 Comparing Utility Loss -- 6 Summary, Open Issues, Conclusions -- Appendix -- A.1 Approximate Behavior of Utility Loss Indicator U for CKM in Extremely Detailed Tabulations -- Appendix A.2 -- Appendix A.3 -- References -- On Privacy of Multidimensional Data Against Aggregate Knowledge Attacks -- 1 Introduction -- 2 Related Work -- 3 Problem Statement -- 3.1 Preliminaries -- 3.2 Problem Definition -- 4 Privacy-Preserving Method -- 4.1 Preprocessing Step -- 4.2 Space Allocation Step -- 4.3 View Creation Step -- 5 Experimental Evaluation -- 6 Discussion -- 7 Conclusion -- References -- Synthetic Decimal Numbers as a Flexible Tool for Suppression of Post-published Tabular Data -- 1 Introduction -- 2 A Motivating Example -- 3 A Theoretical Framework -- 4 Application of the Theory -- 4.1 Application to Frequency Tables -- 4.2 Precision and Implementation -- 4.3 Application to Magnitude Tables -- 4.4 Detection of Disclosure Risk -- 5 Real Applications -- 5.1 Register-Based Employment Statistics -- 5.2 Commissioned Data in Business Statistics -- 6 Concluding Remarks -- References -- Disclosure Risk Assessment and Record Linkage.
The Risk of Disclosure When Reporting Commonly Used Univariate Statistics -- 1 Introduction -- 2 Methodology -- 3 Results -- 4 Discussion Example -- 5 Graphical Representation -- 6 Conclusion -- Appendix -- References -- Privacy-Preserving Protocols -- Tit-for-Tat Disclosure of a Binding Sequence of User Analyses in Safe Data Access Centers -- 1 Introduction -- 2 Background -- 3 Binding to a Sequence of Analyses with Tit-for-Tat Analysis Disclosure -- 3.1 Instruction Preparation -- 3.2 Hash-Based Coding and Execution -- 3.3 Security Properties -- 4 Using Binding Sequences of User Analyses to Ensure Ethical Compliance -- 4.1 Checking Confidentiality -- 4.2 Ex ante Checking of Explainability and Fairness -- 5 Conclusions and Further Research -- References -- Secure and Non-interactive k-NN Classifier Using Symmetric Fully Homomorphic Encryption -- 1 Introduction -- 2 Background -- 2.1 Homomorphic Encryption -- 2.2 Functional Bootstrap in TFHE -- 3 Our Contribution -- 3.1 The System Model -- 3.2 Encrypted k-NN Challenges -- 3.3 Our Proposed k-NN Algorithm -- 4 Performance Evaluation -- 4.1 Test Environment -- 4.2 Performance Results -- 5 Conclusion -- References -- Unstructured and Mobility Data -- Automatic Evaluation of Disclosure Risks of Text Anonymization Methods -- 1 Introduction -- 2 Background -- 3 A Re-identification Attack for Evaluating Anonymized Text -- 3.1 Building the Classifier -- 4 Empirical Experiments -- 4.1 Results -- 5

Conclusions and Future Work -- References -- Generation of Synthetic Trajectory Microdata from Language Models -- 1 Introduction -- 2 Related Work -- 2.1 Sequential Models for Trajectory Prediction -- 2.2 Privacy-Preserving Tajectory Data Publishing -- 3 Synthetic Trajectory Generation Method -- 3.1 Data Preprocessing -- 3.2 Next-Point Prediction Model -- 3.3 Synthetic Data Generation -- 4 Experimental Analysis.
4.1 Data Sets and Preprocessing -- 4.2 Model Training -- 4.3 Results of Data Generation -- 4.4 Additional Remarks on the Experimental Work -- 5 Conclusions and Future Work -- References -- Synthetic Data -- Synthetic Individual Income Tax Data: Methodology, Utility, and Privacy Implications -- 1 Introduction -- 2 Data Synthesis Methodology -- 2.1 Administrative Tax Data -- 2.2 Synthetic Data Generation -- 2.3 Disclosure Risk Measures -- 3 Evaluation -- 4 Conclusions and Future Work -- References -- On Integrating the Number of Synthetic Data Sets m into the a priori Synthesis Approach -- 1 Introduction -- 2 Review of the Use of Saturated Models for Synthesis -- 2.1 The Metrics -- 3 The Role of m as a Tuning Parameter -- 3.1 Obtaining Inferences from m&gt -- 1 Data Sets -- 4 Introducing the 3(k,d) and 4(k,d) Metrics -- 5 Empirical Study -- 5.1 Measuring Risk -- 5.2 Measuring Utility -- 5.3 Tuning m and in Relation to the Risk-Utility Trade-Off -- 6 Discussion -- References -- Challenges in Measuring Utility for Fully Synthetic Data -- 1 Introduction -- 2 Measuring the Utility -- 2.1 A Global Utility Measure: The pMSE -- 2.2 Two Outcome-Specific Measure: The Confidence Interval Overlap and the Mean Absolute Standardized Coefficient Difference -- 3 Misleading Utility Measures: An Illustration -- 3.1 Synthesis Strategies -- 3.2 Results for the Fit-for-Purpose Measures -- 3.3 Results for the Outcome Specific Measures -- 3.4 Results for the Global Utility Measures -- 4 Conclusions -- References -- Comparing the Utility and Disclosure Risk of Synthetic Data with Samples of Microdata -- 1 Introduction -- 2 Background -- 2.1 Data Synthesis -- 2.2 Synthetic Census Microdata -- 3 Research Design -- 3.1 Data Synthesisers -- 3.2 Data -- 3.3 Measuring Disclosure Risk Using TCAP -- 3.4 Evaluating Utility -- 4 Results -- 5 Discussion -- 6 Conclusion -- References.
Utility and Disclosure Risk for Differentially Private Synthetic Categorical Data -- 1 Introduction -- 2 Methods for Creating Synthetic Data Sets -- 2.1 Without DP Guarantee -- 2.2 Adapting Methods for Synthetic Data to Make Them DP -- 3 Measures of Utility and Disclosure Risk for Synthetic Categorical Data -- 3.1 Disclosure Risk -- 3.2 Utility -- 4 Data Sets Used for the Evaluation -- 5 Results -- 5.1 Utility and Disclosure Risk for Non-DP Synthesis -- 5.2 Utility and Disclosure Risk for DP Synthesis -- 6 Discussion and Future Work -- A Appendix -- A. 1 Details of the Variables in Data Sets -- References -- Machine Learning and Privacy -- Membership Inference Attack Against Principal Component Analysis -- 1 Introduction -- 2 Background -- 2.1 Principal Component Analysis -- 2.2 Membership Inference Attacks -- 3 Related Work -- 4 Membership Inference Attacks Against PCA -- 4.1 Threat Model and Attack Methodology -- 4.2 Experimental Setup -- 4.3 Experimental Results -- 5 Differentially-Private PCA and MIA -- 5.1 Preliminaries on Differential Privacy -- 5.2 Differentially Private PCA Approaches -- 5.3 Experimental Results -- 6 Conclusion -- References -- When Machine Learning Models Leak: An Exploration of Synthetic Training Data -- 1 Introduction -- 2 Threat Model -- 3 Background and Related Work -- 3.1 Propensity to Move -- 3.2 Privacy in Machine Learning -- 3.3 Attribute Inference Attack -- 4 Experimental Setup -- 4.1 Data Set -- 4.2 Utility Measures -- 4.3 Adversary Resources -- 5 Experimental Results -- 5.1 Evaluation of Machine Learning Algorithms

-- 5.2 Model Inversion Attribute Inference Attack -- 6 Conclusion and Future Work -- References -- Case Studies -- A Note on the Misinterpretation of the US Census Re-identification Attack -- 1 Introduction -- 2 The Bureau's Re-identification Attack -- 3 Our Simple Inference Non-attack.
3.1 Effect of Majority Race/Ethnicity Threshold.

Sommario/riassunto

This book constitutes the refereed proceedings of the International Conference on Privacy in Statistical Databases, PSD 2022, held in Paris, France, during September 21-23, 2022.The 25 papers presented in this volume were carefully reviewed and selected from 45 submissions. They were organized in topical sections as follows: Privacy models; tabular data; disclosure risk assessment and record linkage; privacy-preserving protocols; unstructured and mobility data; synthetic data; machine learning and privacy; and case studies.