People's Democratic Republic of Algeria
Ministry of Higher Education and Scientific Research
Ferhat Abbas University Setif 1
Faculty of Sciences
Computer Science Department



Ferhat Abbas University Setif 1

A Master's thesis submitted to the Faculty of Sciences for
obtaining the degree of **Master in Computer Science**
Specialization: Data Engineering and Web Technology

Theme

Privacy Protection in the Era of Big Data and Artificial Intelligence

---

Realized by:

Zitouni Taha Abdelmoudjib
Djelaoudji Abdelouahab

Supervised by:

Dr. Hadi Fairouz

University Year: 2025–2026

Abstract

The rapid growth of Big Data and Artificial Intelligence has significantly increased the collection and analysis of personal data, raising important privacy concerns including re-identification, membership inference, and model inversion attacks. Single-technique approaches such as anonymisation or Differential Privacy applied independently offer limited protection against adversaries possessing auxiliary background knowledge. This thesis investigates a hybrid approach combining k-Anonymity and Differential Privacy (ε-DP) on the UCI Adult Income dataset, where k-Anonymity first reduces re-identification risks through quasi-identifier generalisation and equivalence class formation, followed by Differential Privacy mechanisms based on Laplace noise and Randomised Response with varying privacy budgets. A LinearSVC classifier is trained and evaluated using accuracy, precision, recall, and F1-score.

k-Anonymity ($k = 5$) achieves F1 = 64.20% versus 53.58% for the unprotected baseline, confirmed by 5-fold cross-validation, and attributable to feature space compression and implicit regularisation rather than data leakage. Differential Privacy alone substantially reduces model utility at strong privacy budgets, while the hybrid approach outperforms it: at $\varepsilon = 1.0$, the hybrid reaches F1 = 58.33% compared to 32.43% for Differential Privacy on raw data, suggesting that k-Anonymity preprocessing reduces noise sensitivity and improves model stability. The study highlights the potential benefits of combining structural anonymisation with statistical privacy mechanisms in privacy-preserving machine learning, while discussing the limitations of the proposed approach and the trade-off between privacy protection and model utility.

Keywords: privacy-preserving machine learning, k-anonymity, differential privacy, re-identification, membership inference attack, GDPR, Big Data, artificial intelligence.

Résumé

La croissance rapide du Big Data et de l'Intelligence Artificielle a accru la col-lecte et l'analyse des données personnelles, soulevant des préoccupations de vie privée telles que les attaques de ré-identification, d'inférence d'appartenance et d'inversion de modèle. Les approches reposant sur une seule technique de protection, comme l'anonymisation ou la confidentialité différentielle appliquées indépendamment, offrent une protection limitée contre des adversaires disposant de connaissances auxiliaires. Ce mémoire étudie une approche hybride combinant le k-anonymat et la confidentialité différentielle (ε-DP) sur le jeu de données UCI Adult Income, où le k-anonymat réduit d'abord les risques de ré-identification par la généralisation des quasi-identifiants et la formation de classes d'équivalence, suivi de mécanismes de confidentialité différentielle fondés sur le bruit de Laplace et la réponse aléatoire. Un classificateur LinearSVC est évalué selon la précision, le rappel et le score F1.

Le k-anonymat ($k = 5$) atteint F1 = 64,20% contre 53,58% pour la référence non protégée, confirmé par une validation croisée à 5 plis ; ce résultat est attribuable à la compression de l'espace des caractéristiques et à une régularisation implicite, et non à une fuite de données. La confidentialité différentielle seule réduit considérablement l'utilité à des budgets stricts, tandis que l'approche hybride la surpasse : pour $\varepsilon = 1{,}0$, le modèle hybride atteint F1 = 58,33% contre 32,43% pour la confidentialité différentielle seule, suggérant que le prétraitement par k-anonymat réduit la sensibilité au bruit et améliore la stabilité du modèle. L'étude met en évidence les avantages de la combinaison de mécanismes d'anonymisation structurelle et de protection statistique de la vie privée, tout en discutant des limitations de l'approche proposée et du compromis entre protection de la vie privée et utilité du modèle.

Mots-clés : apprentissage automatique préservant la vie privée, k-anonymat, confidentialité différentielle, ré-identification, attaque par inférence d'appartenance, RGPD, Big Data, intelligence artificielle.

الملخص

أسفر النمو السريع للبيانات الضخمة والذكاء الاصطناعي عن زيادة ملحوظة في جمع البيانات الشخصية وتحليلها، مثيرًا مخاوف تتعلق بالخصوصية كهجمات إعادة التعرّف على الأفراد واستدلال العضوية وعكس النموذج. وتوفر
الأساليب المعتمدة على تقنية واحدة، كإخفاء الهوية أو الخصوصية التفاضلية عند تطبيق كل منهما بشكل مستقل، حماية محدودة في مواجهة الخصوم الذين يمتلكون معرفة خلفية إضافية. تستكشف هذه المذكرة نهجًا هجينًا يجمع بين إخفاء
الهوية k والخصوصية التفاضلية (ε-DP) على مجموعة بيانات UCI Adult Income، إذ يُطبَّق إخفاء الهوية k أولًا للحد من مخاطر إعادة التعرّف عبر تعميم المعرّفات شبه الفريدة وتكوين فئات التكافؤ، تليه آليات الخصوصية التفاضلية
المعتمدة على ضوضاء لابلاس والاستجابة العشوائية بميزانيات خصوصية متفاوتة. يُدرَّب مصنف LinearSVC ويُقيَّم باستخدام مقاييس الدقة والاستدعاء ودرجة F_1.

حقَّق إخفاء الهوية k ($k = 5$) درجة F_1 بلغت 64،20% مقابل 53،58% للنموذج الأساسي غير المحمي، وأُكِّد ذلك بالتحقق المتقاطع بخمس طيات، ويُعزى هذا التحسن إلى ضغط فضاء الميزات والتنظيم الضمني لا إلى أي تسرب
في البيانات. تُقلص الخصوصية التفاضلية وحدها أداء النموذج بشكل ملحوظ عند الميزانيات الصارمة، في حين يتفوق النهج الهجين عليها: عند $\varepsilon = 1,0$، بلغت درجة F_1 للنموذج الهجين 58،33% مقابل 32،43% للخصوصية التفاضلية وحدها، مما يشُير إلى أن المعالجة المسبقة بـ إخفاء الهوية k تُقلص الحساسية للضوضاء وتُحسّن استقرار النموذج. تُبرز الدراسة فوائد الجمع بين آليات إخفاء الهوية البنيوية وحماية الخصوصية الإحصائية في أنظمة التعلم الآلي، مع مناقشة
قيود النهج المقترح والمقايضة بين حماية الخصوصية وأداء النموذج.

الكلمات المفتاحية: التعلم الآلي المحافظ على الخصوصية، إخفاء الهوية k، الخصوصية التفاضلية، إعادة التعرّف على الأفراد، هجوم استدلال العضوية، اللائحة العامة لحماية البيانات (GDPR)، البيانات الضخمة، الذكاء الاصطناعي.

Contents

Abstract 2

Résumé 3

الملخص 4

List of Figures 10

List of Tables 12

List of Abbreviations 13

General Introduction 15

1 Technical Foundations 17

1.1 Introduction 17

1.2 Big Data and Artificial Intelligence 17

1.2.1 Data Pipelines in AI Systems 17

1.2.2 Machine Learning Models 18

1.2.3 Data Lifecycle and Privacy Exposure 19

1.3 Technical Threats to Privacy in AI Systems 19

1.3.1 Re-identification Attacks 19

1.3.2 Membership Inference Attacks 19

1.3.3 Model Inversion Attacks 20

1.4 Privacy-Enhancing Technologies 20

1.4.1 Anonymisation and *k*-Anonymity 20

1.4.2 Differential Privacy 21

1.4.3 Federated Learning and Encryption-Based Approaches 22

1.5 Comparative Analysis of Protection Techniques 22

1.6 Conclusion 24

2 Problem Analysis and System Design 25

2.1 Introduction 25

2.2 Related Work and Critical Analysis 25

2.2.1 Differential Privacy in Machine Learning Systems 25

2.2.2 Anonymisation and PII Detection Approaches 26
2.2.3 Federated Learning for Distributed Privacy Preservation 27
2.2.4 Generative AI and Advanced Privacy Mechanisms 27
2.2.5 Comparative Summary of Related Works 28
2.3 Research Gap and Positioning 28
2.4 Problem Description 30
2.4.1 Data Type 30
2.4.2 Use Case and Privacy Challenge 31
2.5 Experimental Assumptions 31
2.6 Proposed System Architecture 31
2.7 Design Rationale 32
2.8 Conclusion 34

3 Implementation of the Proposed System 35
3.1 Introduction 35
3.2 Development Environment 36
3.2.1 Development and Execution Environment 36
3.2.2 Tools and Libraries 36
3.3 Dataset Preparation 37
3.3.1 Loading and Cleaning 37
3.3.2 Encoding Categorical Variables 37
3.3.3 Train/Test Split 37
3.3.4 Feature Normalisation 38
3.4 Baseline Model Implementation 38
3.4.1 Principle of LinearSVC 38
3.4.2 Parameters and Justification 39
3.4.3 Role of the Baseline 39
3.5 Implementation of k-Anonymity 40
3.5.1 Quasi-Identifier Selection 40
3.5.2 Generalisation Strategy 41
3.5.3 Concrete Generalisation Example 41
3.5.4 Suppression and k Value Selection 42
3.6 Implementation of Differential Privacy 43
3.6.1 Definition and Privacy Guarantee 43
3.6.2 Laplace Mechanism for Numerical Columns 43
3.6.3 Randomised Response for Categorical Columns 45
3.6.4 Choice of ε Values 45
3.7 Parameter Selection and Justification 45
3.7.1 Justification of k Values 46

3.7.2 Justification of ε Values 46
3.7.3 Justification of Differential Privacy Sensitivity Values 46
3.7.4 Justification of Classifier Choice 47
3.7.5 Justification of Dataset Choice 47
3.7.6 Pseudo-Code of the Hybrid Pipeline 48
3.8 Implementation of the Hybrid Architecture 48
3.8.1 Integration of the Two Mechanisms 48
3.8.2 Order of Processing 49
3.8.3 Module Interaction 49
3.9 Experimental Configuration 51
3.9.1 Parameter Values 51
3.9.2 Evaluation Metrics 51
3.9.3 Validation Protocol 52
3.10 Conclusion 52

4 Experimental Evaluation and Discussion 54
4.1 Introduction 54
4.2 Experimental Setup 55
4.2.1 Hardware and Software Configuration 55
4.2.2 Dataset 55
4.2.3 Train/Test Split and Cross-Validation 55
4.2.4 Parameter Values Tested 55
4.2.5 Evaluation Metrics 55
4.3 Results and Analysis 56
4.3.1 Global Performance Comparison 56
4.3.2 Analysis of k-Anonymity 59
4.3.3 Analysis of Differential Privacy 62
4.3.4 Analysis of the Hybrid Architecture 63
4.3.5 Cross-Validation Stability Analysis 65
4.4 Discussion and Comparison with Related Work 67
4.4.1 Comparison with Related Works 67
4.4.2 Consistency with the Literature 69
4.4.3 Differences and Open Questions 70
4.5 Threats to Validity 70
4.5.1 Threats Related to the Dataset 70
4.5.2 Threats Related to the Model 70
4.5.3 Threats Related to the Privacy Evaluation 71
4.5.4 Threats Related to Generalisation 71
4.6 Theoretical Privacy Assessment 71

4.6.1 Re-identification Attack 71
4.6.2 Membership Inference Attack 72
4.6.3 Model Inversion Attack 72
4.6.4 Summary of Theoretical Assessment 72
4.7 Conclusion 73

General Conclusion **74**

Bibliography **77**