PROGRAMA DE CURSO
PRIVACIDAD DE DATOS

A. Antecedentes generales del curso:

Departamento	Ciencias de la Computación					
Nombre del curso	Privacidad de datos	Código	CC5215	Créditos	6	
Nombre del curso en inglés	*Data privacy*					
Horas semanales	Docencia	3	Auxiliares	1,5	Trabajo personal	5,5
Carácter del curso	Obligatorio		Electivo especialidad	X		
Requisitos	MA3403,CC3201,(CC5205/AUTOR)					

B. Propósito del curso:

El propósito de este curso es que los y las estudiantes aprendan a analizar datos que involucren información sensible y/o confidencial sobre individuos, preservando la privacidad de los mismos. En particular, al finalizar el curso los y las estudiantes serán capaces, por un lado, de reconocer distintos escenarios en el análisis de datos que pueden dar lugar a la filtración de información sensible, y por otro lado, a aplicar técnicas que permitan realizar el análisis de manera segura.

Se partirá estudiando técnicas tradicionales de anonimización, destinadas a liberar microdatos sobre los individuos. Luego se estudiará la privacidad diferencial, técnica introducida más

recientemente, y que resuelve el problema de la privacidad de manera más satisfactoria. A lo largo del curso cubriremos tanto la práctica como la teoría de la privacidad de datos.

Las competencias específicas (CE) y genéricas (CG) a las que tributa el curso son:

CE2: Analizar, diseñar y/o adaptar algoritmos y estructuras de datos que cumplan con las garantías requeridas de correctitud y eficiencia.

CE3: Gestionar bases de datos utilizando modelos, lenguajes de consulta asociados, técnicas eficientes de acceso a datos y aplicación de políticas de seguridad, con la finalidad de obtener información relevante.

CG1: Comunicación académica y profesional

Comunicar en español de forma estratégica, clara y eficaz, tanto en modalidad oral como escrita, puntos de vista, propuestas de proyectos y resultados de investigación fundamentados, en situaciones de comunicación compleja, en ambientes sociales, académicos y profesionales.

CG2: Comunicación en inglés

Leer y escuchar de manera comprensiva en inglés variados tipos de textos e informaciones sobre temas concretos o abstractos, comunicando experiencias y opiniones, adecuándose a diferentes contextos de acuerdo a las características de la audiencia.

CG3: Compromiso ético

Actuar de manera responsable y honesta, dando cuenta en forma crítica de sus propias acciones y sus consecuencias, en el marco del respeto hacia la dignidad de las personas y el cuidado del medio social, cultural y natural.

CG4: Trabajo en equipo

Trabajar en equipo, de forma estratégica y colaborativa, en diversas actividades formativas, a partir de la autogestión de sí mismo y de la relación con el otro, interactuando con los demás en diversos roles: de líder, colaborador u otros, según requerimientos u objetivos del trabajo, sin discriminar por género u otra razón.

C. Resultados de aprendizaje:

Competencias específicas	Resultados de aprendizaje
CE2	RA1: Adapta algoritmos de análisis de datos sobre individuos, a fin de que los resultados de los análisis resguarden información sensible de los individuos.
	RA2: Ejecuta un análisis formal de algoritmos de análisis de datos, a fin de cuantificar formalmente estos algoritmos en base a sus niveles de privacidad y utilidad.
CE3	RA3: Adapta y/o corrige malas prácticas en el manejo de datos sensibles, identificando este tipo de malas prácticas en base al tipo de ataque sobre ese manejo de datos.
	RA4: Diseña políticas de acceso a datos sensibles sobre individuos, con el objetivo de preservar su privacidad, considerando la forma o tipo de dato al que se accederá y de la frecuencia con la que se requiere acceder a dichos datos.
Competencias genéricas	**Resultados de aprendizaje**
CG1	RA5: Comunica en forma oral y escrita, sobre distintos aspectos del manejo de datos confidenciales, haciendo uso del lenguaje técnico, para establecer un lenguaje común y colaborar con sus pares en las tareas asignadas.
CG2	RA6: Lee en inglés, de manera analítica y comprensiva, artículos científicos y libros de texto sobre privacidad de datos a fin de generar nuevos conocimientos atingentes y aplicables al análisis de datos sensibles.
CG3	RA7: Cumple obligaciones y acuerdos, respetando los compromisos adquiridos, reflexionando sobre sus acciones y asumiendo las consecuencias.
CG4	RA8: Trabaja con su equipo, de forma colaborativa y organizada, a fin de contribuir con su labor al desarrollo y logro del proyecto de base de datos.

D. Unidades temáticas:

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
1	RA3	Introducción al análisis de datos sensibles	1 semana
Contenidos		**Indicador de logro**	
1.1. El problema del análisis de datos sensibles 1.2. Introducción a las técnicas de anonimización y privacidad diferencial.		El/la estudiante: 1. Reconoce las causas que originan la filtración de datos sensibles. 2. Reconoce los desafíos (conflicto "privacidad de individuos v/s utilidad de los datos") que involucran el análisis de datos sensibles. 3. Clasifica la noción de privacidad en base a un tipo de interpretación que la considera como una propiedad de los datos *per se*, o como una propiedad de los algoritmos que generan dichos datos. 4. Identifica dos estrategias para abordar el análisis de datos sensibles: anonimización y privacidad diferencial. 5. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 6. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 7. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.	
Bibliografía de la unidad		[1] Cap. 1 [2] Cap. 1 [4] Cap. 1	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
2	RA1-RA2-RA3-RA4-RA5-RA6-RA7	Anonimización	3 semanas
Contenidos		**Indicador de logro**	

2.1. Ataques por asociación de registro, asociación de atributo y asociación de tabla. 2.2. Nociones de *k*-anonimidad, *k*-(X,Y)-anonimidad y *l*-diversidad (distinta y entrópica). 2.3. Operaciones de generalización, enmascarado, eliminación y nulificación, anatomización y perturbación aleatoria. 2.4 Métricas de distorsión mínima y pérdida de la información. 2.5 Privacidad para la publicación de múltiples tablas.	El/la estudiante: 1. Identifica diferentes modelos de ataques a la privacidad. 2. Clasifica distintas nociones de privacidad, de acuerdo con el tipo de que se busca evitar. 3. Reconoce y aplica diferentes operaciones de anonimización para establecer nociones de privacidad. 4. Identifica y utiliza diferentes métricas para evaluar, mediante un análisis formal, la precisión de los procesos de anonimización. 5. Identifica y analiza los desafíos nuevos que están involucrados en el resguardo de la privacidad y cómo abordarlos, considerando el uso de una única tabla versus el uso de múltiples tablas. 6. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 7. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 8. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.
Bibliografía de la unidad	[1] Cap. 2, 3, 4 y 8. [4] Cap. 2 y 3.

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
3	RA1-RA2-RA3-RA4-RA5-RA6-RA7	Introducción a la privacidad diferencial	3 semanas
Contenidos		Indicador de logro	
3.1. Motivación y definición de la noción de privacidad diferencial. 3.2. Teoremas de posprocesamiento, composición secuencial y composición paralela. 3.3. Mecanismos de Laplace y Exponencial. 3.4 Privacidad de grupos.		El/la estudiante: 1. Reconoce las limitaciones de las técnicas de anonimización y las restricciones impuestas por la ley fundamental de la recuperación de la información. 2. Describe la noción de privacidad diferencial, reconociendo y explotando sus beneficios. 3. Diseña algoritmos diferencialmente privados para análisis de datos numéricos y categóricos, adaptando sus versiones clásicas a través de los	

3.5 Sensibilidad y clipping.	mecanismos de Laplace y Exponencial, respectivamente. 4. Calcula la sensibilidad global de análisis de datos numéricos, recurriendo a técnicas de clipping en el caso de análisis de resultados no acotados. 5. Determina, a través de un razonamiento formal, los niveles de privacidad y utilidad ofrecidos por algoritmos diferencialmente privados. 6. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 7. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 8. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.
Bibliografía de la unidad	[2] Cap. 2 y 3. [3] Cap. 1 y 2. [4] Cap. 4, 5, 6 y 10.

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
4	RA1-RA2-RA3-RA4 -RA5-RA6-RA7	Variantes de privacidad diferencial y mecanismos avanzados de composición	1,5 semanas
Contenidos		Indicador de logro	
4.1. Privacidad diferencial aproximada, de Rényi y cero-concentrada. 4.2 Teorema de composición *k-fold.*		El/la estudiante: 1. Reconoce y compara distintas variantes de la noción de privacidad diferencial junto a sus respectivas propiedades, identificando las ventajas y desventajas de cada variante. 1. Aplica mecanismos avanzados de composición para diseñar algoritmos diferencialmente privados con mejores propiedades de privacidad.	

	2. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 3. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 4. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.
Bibliografía de la unidad	[4] Cap. 7 y 9

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
5	RA1-RA2-RA3-RA4-RA5-RA6-RA7	Sensibilidad local	2 semana
Contenidos		Indicador de logro	
5.1. Sensibilidad suave, *propose-test-release*, liberación de valores estables y cota diferencialmente privada de la sensibilidad local.		El/la estudiante: 1. Identifica las limitaciones de las técnicas de privacidad diferencial basadas en la sensibilidad global. 2. Diseña y analiza algoritmos diferencialmente privados utilizando técnicas basadas en la sensibilidad local. 3. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 4. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 5. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.	
Bibliografía de la unidad		[4] Cap. 8. [3] Cap. 3.	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
6	RA1-RA2-RA3-RA4-RA5-RA6-RA7	Correlación de ruido	2,5 semana
Contenidos		Indicador de logro	
6.1. Algoritmo SmallDB 6.2 Algoritmo MWEM		El/la estudiante: 1. Utiliza la correlación de ruido para mejorar la precisión de algoritmos diferencialmente privados. 2. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 3. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 4. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.	
Bibliografía de la unidad		[2] Cap. 4. [3] Cap. 4.	

Número	RA al que tributa	Nombre de la unidad	Duración en semanas
7	RA1-RA2-RA3-RA4-RA5-RA6-RA7	Generación de datos sintéticos	2 semana
Contenidos		Indicador de logro	
7.1. Construcción de sinopsis de datos a partir de marginales 1-*way* y *n*-*way*. 7.2 Clustering vía *k-means* diferencialmente privado.		El estudiante: 1. Utiliza algoritmos diferencialmente privados para la generación de datos sintéticos. 2. Cumple obligaciones y acuerdos, respetando los compromisos adquiridos en sus actividades académicas. 3. Planifica y entrega sus tareas, basándose en sus capacidades, sin incurrir en plagio, copia, suplantación de identidad. 4. Lee de manera comprensiva diversas fuentes en inglés sobre privacidad de datos, determinando sus ideas principales.	
Bibliografía de la unidad		[4] Cap. 15. [5]	

E. Estrategias de enseñanza:

El curso se estructura en base a distintas metodologías que incluyen principalmente:

- **Clases expositivas** que combinan el uso de diapositivas, la pizarra, y la programación en vivo para presentar el material.
- **Laboratorios**, donde los y las estudiantes identifican los problemas fundamentales en el análisis de datos sensibles, así como modelos y técnicas para abordarlo.
- **Proyecto final**, donde los y las estudiantes identifican un problema de privacidad de datos en un análisis de la vida real, para luego diseñar, implementar y evaluar una herramienta que haga el análisis protegiendo la privacidad.

E. Estrategias de evaluación:

El curso plantea las siguientes instancias de evaluación:

Tipo de evaluación	Resultado de aprendizaje asociado a la evaluación
Laboratorios (8 a 12) que podrán ser individuales o grupales (como especificado por el profesor o la profesora a cargo al principio del semestre) y se promediaron en partes iguales para obtener la nota de laboratorio (NL).	Evalúa RA1, RA2, RA3, RA4, RA5, RA6 y R7.
Proyecto final (1) que se realizará en grupos. Se evaluará a través de la entrega de un informe final y una presentación oral, que colectivamente determinarán la nota de proyecto (NP). El desarrollo del proyecto es transversal a todas las unidades y busca diseñar políticas de acceso a datos sensibles	Evalúa RA3, RA4, RA5, RA6, RA7 y RA8.

sobre individuos, utilizando distintas herramientas conceptuales y algorítmicas.	

Para aprobar el curso es necesario que NL $\geq$ 4 y NP $\geq$ 4, y en dicho caso la nota final (NF) se calculará como

NF = 0,65 * NL + 0,35 * NP.

Si NL $<$ 4 o NP $<$ 4 la nota final se calculará como

NF = min {NL, NP}.

F. Recursos bibliográficos:

Bibliografía obligatoria:

[1] *Introduction to privacy-preserving data publishing concepts and techniques.* Fung, B. C., Wang, K., Fu, A. W. C., & Philip, S. Y., Chapman and Hall/CRC. 2010.

[2] *The algorithmic foundations of differential privacy.* Foundations and Trends in Theoretical Computer Science: Vol. 9: No. 3–4, pp 211-407. Dwork, C. & Roth, A. 2014. Disponible online: https://www.nowpublishers.com/article/Details/TCS-042

[3] *The complexity of differential privacy.* Tutorials on the Foundations of Cryptography, pp. 347-450. Vadhan, S. Springer, Cham. 2017. Disponible online: https://salil.seas.harvard.edu/publications/complexity-differential-privacy

[4] *Programming Differential Privacy*. Near, J. P. & Abuah, C. 2011. Disponible online: https://programming-dp.com/

[5] *Differentially private k-means clustering*. Su, D., Cao, J., Li, N., Bertino, E. & Jin, H. En *Proceedings of the 6th ACM Conference on Data and Application Security and Privacy,* pp. 26-37. 2016.

H. Datos generales sobre elaboración y vigencia del programa de curso:

Vigencia desde:	Otoño 2023
Elaborado por:	Federico Olmedo y Matías Toro
Validado por:	Validado por académico par: Aidan Hogan Validado por CTD de Computación
Revisado por:	Área de Gestión Curricular