



LIBRO DE ACTAS

XXV Workshop de Investigadores en Ciencias de la Computación

Buenos Aires, abril del 2023
UNNOBA – Red de Universidades
de Carreras de Informática RedUNCI



ITT

Instituto
Transferencia
Tecnología



ESCUELA DE
TECNOLOGÍA



UNNOBA



UNIVERSIDAD NACIONAL
NOROESTE • BUENOS AIRES

Datos de la edición

Escuela de Tecnología

Directora: Sarobe, Mónica

Secretario Académico: Cicerchia, Lucas Benjamin

Instituto de Investigación y Transferencia en Tecnología – ITT

Director: Ramón, Hugo

Investigador: Esnaola, Leonardo

Investigador: Tessore, Juan Pablo

Red de Universidades de Carreras de Informática – RedUNCI

Representante: Russo, Claudia

Edita

CEDi Centro de Edición y Diseño. UNNOBA

Callao 289 3.º piso, CP.1022

Tel: 54 -11-535-31520. Ciudad Autónoma
de Buenos Aires, Argentina

Diseño y diagramación de tapa e interior:

CEDi | cedi@unnoba.edu.ar

Fecha de catalogación:

Queda hecho el depósito que marca la ley 11723
Reservados todos los derechos

Abril | 2023

► Diseño : Reflexiones sobre el hacer en el territorio, 2018- 2020 / Maria de las Mercedes Filpe ... [et al.] ; compilación de Alejandra Mizrahi ... [et al.]. - 1a ed. - Junín : Universidad Nacional del Noroeste de la Provincia de Buenos Aires. UNNOBA, 2022.
Libro digital, PDF - (Investigación)

Archivo Digital: descarga

ISBN 978-987-3724-64-0

1. Diseño Indumentaria. 2. Diseño Textil. 3. Educación. I. Filpe, Maria de las Mercedes. II. Alejandra Mizrahi, comp.
CDD 746.92



Esta obra está bajo una
licencia de Creative Commons
Reconocimiento - NoComercial -
SinObraDerivada 4.0 Internacional.

Índice

ARSO – Arquitectura, Redes y Sistemas Operativos	Pág (12-57)
Desafíos y buenas prácticas para pequeños proveedores de Internet zonales.	Pág (13-16)
Migración controlada de procesos en sistemas distribuidos.	Pág (17-21)
Estudios del aporte de tecnologías como BLOCKCHAIN, NFT y VPN a la seguridad en plataformas de IOT.	Pág (22-26)
IoT for Vineyard Posts.	Pág (27-31)
IoT-based energy monitoring system, to deploy an AMI.	Pág (32-36)
In a Micro Smart Grid environment.	Pág (37-41)
Utilización Eficiente y Sostenible de Recursos en Clúster de bajo costo.	Pág (42-46)
Tecnologías de Smart IoT para la solución de problemas en el medio socio productivo.	Pág (47-52)
Despliegue de aplicaciones contenerizadas: un caso de implementación basado en crane.	Pág (53-57)
ASI – Agentes y Sistemas Inteligentes	Pág (58-156)
Interfaces de Programación de Aplicaciones orientadas a la Integración de Sistemas Judiciales.	Pág (59-63)
Organizaciones inteligentes asistidas por tecnologías 4.o.	Pág (64-68)
Aprendizaje Automático para Nodos Inteligentes.	Pág (69-71)
Depuración de bases de datos de segunda sección del Boletín oficial de la república argentina mediante Aprendizaje de máquina.	Pág (72-76)
Una herramienta de Inteligencia Artificial para el modelado de distintos escenarios con la determinación del Índice de Calidad del Agua (ICA) para aplicar a ríos urbanos de Argentina.	Pág (77-81)
Movilidad inteligente en una ciudad sostenible.	Pág (82-86)
Asistente virtual con comportamiento definible de consultas referidas a la legislación argentina para personas con discapacidad.	Pág (87-89)
Automatización robótica de procesos en las pymes.	Pág (90-94)
Clasificación de partículas de arena a través de redes neuronales convolucionales.	Pág (95-100)
Resolución de problemas regionales NP-duros usando técnicas de optimización.	Pág (101-105)
Confección de dataset de patentes de automóviles argentinos para entrenamiento de redes neuronales.	Pág (106-110)
Metaheurísticas aplicadas al alineamiento múltiple de secuencias de ADN y/o de aminoácidos.	Pág (111-114)
Aplicación de inteligencia computacional y computación de alto desempeño en el desarrollo de un modelo de predicción de las condiciones predisponentes al quemado del arroz (<i>pyricularia oryzae</i>).	Pág (115-119)
Lenguaje natural aplicado a los sistemas conversacionales para el aprendizaje de lenguas extranjeras.	Pág (120-124)
Tecnologías Semánticas para el desarrollo de Agentes Inteligentes: Generación de comentarios a partir de código fuente.	Pág (125-128)
Deep Learning para Visión por Computadora.	Pág (129-133)
Inteligencia artificial como herramienta para innovar y dinamizar procesos.	Pág (134-139)
Detección de Ransomware en Blockchains.	Pág (140-142)
Desarrollo de soluciones tecnológicas de bajo costo a problemáticas en el sector hortícola.	Pág (143-147)
Desarrollo de herramientas para generación de material de entrenamiento y análisis de textos estructurados.	Pág (148-152)
Sistema de predicción de incendios forestales.	Pág (153-156)

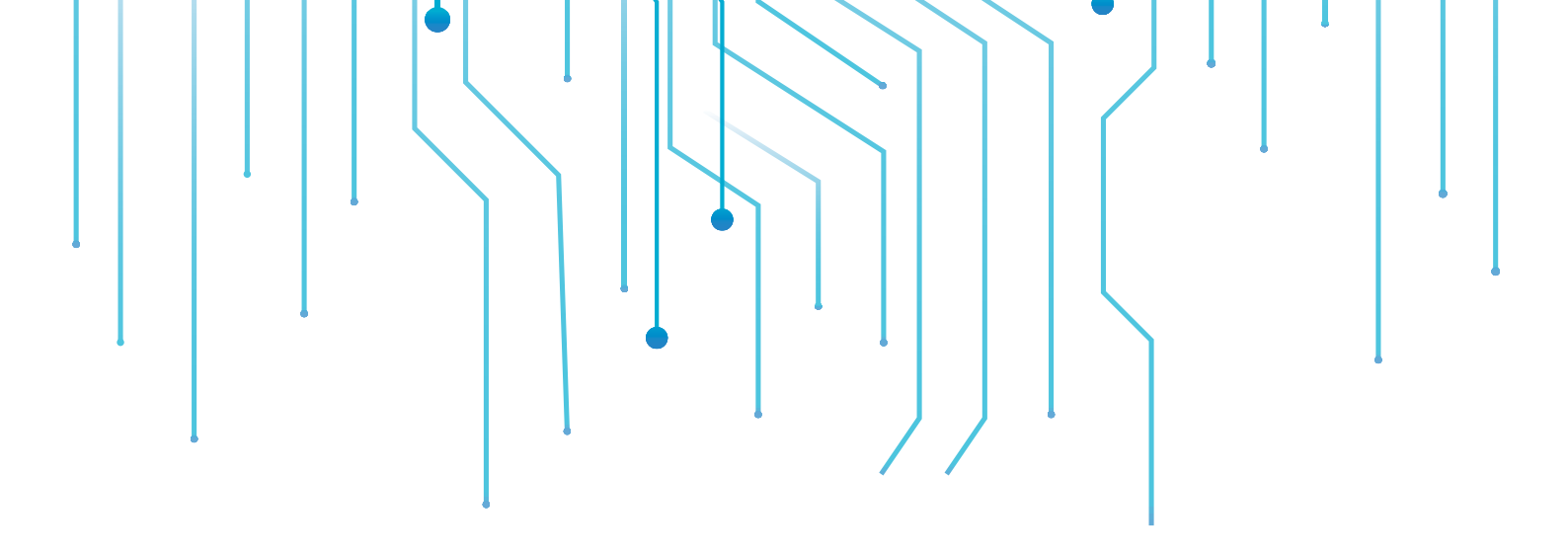
BDDM – Bases de Datos y Minería de Datos	Pág (157-216)
Propuesta de técnicas de validación para la calidad de datos abiertos e identificación de patrones para predicciones con Machine Learning: Segunda Parte.	Pág (158-162)
Knowledge Graphs para accesibilidad y reutilización de activos digitales del Mar Argentino.	Pág (163-166)
Descubrimiento de patrones de comportamiento vinculados al abandono en la Universidad Nacional de Luján mediante la aplicación de técnicas de aprendizaje automático.	Pág (167-173)
Desarrollo de soluciones de inteligencia y analítica de negocios.	Pág (174-177)
Rediseño de la Plataforma Web MultiDB, Cuyo Propósito es el Aprendizaje de Lenguaje SQL, con Acceso a Distintos Motores de Base de Datos.	Pág (178-182)
Sistemas inteligentes en el uso de aplicaciones de bioinformática y sistemas embebidos.	Pág (183-187)
Datos no Estructurados: Indexación, Búsquedas y Aplicaciones	Pág (188-192)
Administración y Recuperación de Información en Bases de Datos Masivas.	Pág (193-197)
Diversas Aplicaciones de Bases de Datos Espaciales y Espacio Temporales.	Pág (198-202)
Bases de Datos no Convencionales.	Pág (203-207)
Comparación de métodos y métricas para la predicción de series temporales aplicados a las ventas masivas de productos minoristas.	Pág (208-212)
Índices de dominio para identificación eficiente mediante código de iris.	Pág (213-216)
CGIV – Computación Gráfica, Imágenes y Visualización	Pág (217-261)
Testing de visualizaciones.	Pág (218-221)
PACS Cloud Open Source.	Pág (222-226)
Análisis visual de movimientos oculares en entornos dinámicos.	Pág (227-231)
Visualización comparativa: estudio y diseño de visualizaciones orientadas a la comparación de conjunto de datos.	Pág (232-236)
Tecnologías inmersivas y visualización situada aplicadas a geociencias.	Pág (237-241)
Visualización de datos multidimensionales procedentes de las geociencias.	Pág (242-246)
Análisis visual guiado para el estudio de datos geoquímicos.	Pág (247-251)
Industrias Inteligentes: Detección de uso de Elementos de Protección Personas en una Planta de Reciclaje.	Pág (252-256)
La importancia de la realidad extendida en el metaverso.	Pág (257-261)
IEI – Innovación en Educación en Informática	Pág (262-331)
Adopción de verilog en materias de diseño digital.	Pág (263-267)
Aplicación con realidad aumentada para el aprendizaje de los conceptos de programación.	Pág (268-272)
Innovación Tecnológica y Metodológica en Educación.	Pág (273-277)
Evaluación de la Calidad de Objetos de Aprendizaje, para la enseñanza de la Informática, desde el Enfoque de los estudiantes. Medición del significado afectivo.	Pág (278-282)
Marco para desarrollar nuevas unidades de proyectos e incorporar docentes y alumnos en actividades de investigación.	Pág (283-287)
Herramientas colaborativas multiplataforma para el aprendizaje basado en competencias en la ingeniería de software.	Pág (288-295)
El portafolio como metodología de trabajo en un grupo de investigación de infraestructura it.	Pág (296-300)
Enseñar programación en la universidad en tiempos de gran demanda de programadores.	Pág (301-305)
Eye tracking para la enseñanza de la programación.	Pág (306-310)
Demanda laboral y competencias en inteligencia artificial dentro de carreras grado en ciencias aplicadas.	Pág (311-315)

Proyecto Esport UNLP Diplomatura en Deportes Electrónicos.	Pág (316-321)
Programación y Autismo: una arquitectura epistémica centrada en Ciencia Ciudadana.	Pág (322-326)
Desarrollo de Recursos Educativos para la Enseñanza de las Ciencias de la Computación en la Escuela Secundaria.	Pág (327-331)
IS – Ingeniería en Software	Pág (332-481)
Principios esenciales de la ingeniería de software en ámbito público.	Pág (333-337)
Catálogo y recomendaciones de recursos (métodos) para facilitar experiencias de Diseño Participativo usando Design Thinking.	Pág (338-342)
Análisis de viabilidad del despliegue de sistemas de software en PyMES.	Pág (343-347)
Métodos, técnicas y herramientas para mejorar y evaluar la calidad de requisitos en proyectos ágiles de software.	Pág (348-352)
Métodos, técnicas y herramientas para mejorar y evaluar la calidad de la capacidad de producción de entregables dentro de un proceso ágil de desarrollo de software	Pág (353-357)
Creación de un observatorio tecnológico de la gestión del conocimiento en la ciudad de buenos aires.	Pág (358-362)
Aprovisionamiento automático en la nube mediante la Ingeniería Dirigida por Modelos.	Pág (363-367)
Hacia la evaluación de la calidad de datos abiertos.	Pág (368-372)
Aplicaciones Móviles, Realidad Virtual y Realidad Aumentada.	Pág (373-376)
Evaluación de las políticas de protección de datos personales en las iniciativas de apertura de datos en el ámbito de la Universidad Nacional de Catamarca.	Pág (377-381)
Desarrollo de Computadoras de a Bordo para Misiones Satelitales de Órbita Baja.	Pág (382-386)
Enfoques y Tendencias en el Desarrollo de Aplicaciones para Dispositivos Móviles.	Pág (387-392)
Construcción de colecciones de proyectos de software para estudios empíricos.	Pág (393-397)
Buenas prácticas de calidad de software en el ámbito público.	Pág (398-402)
Evaluación de lenguajes de modelado gráficos de procesos de negocio colaborativos utilizando un método multicriterio.	Pág (403-407)
Diseño de un marco de trabajo para la gestión del conocimiento de productos de software en evolución.	Pág (408-410)
Aspectos de ingeniería de software, bases de datos relacionales, y bases de datos NOSQL para el desarrollo de sistemas de software híbridos.	Pág (411-417)
Desarrollo de Software Público: Un aporte para la mejora de la calidad.	Pág (418-422)
Cultura, turismo y tecnología. Experiencias turísticas significativas para circuitos históricos de caleta Olivia.	Pág (423-427)
Desarrollo y evaluación de productos web y móvil centrados en la experiencia de usuario.	Pág (428-433)
Evaluación de la Sostenibilidad en un Modelo de Calidad del Software.	Pág (434-438)
CoVaMaT: Modelo conceptual de una herramienta para el soporte a la gestión de variedad en Sistemas Big Data.	Pág (439-443)
Proceso de validación de requerimientos aplicando técnicas de procesamiento de lenguaje natural en un entorno colaborativo.	Pág (444-448)
Computación afectiva aplicada a la interacción con Interfaces Gráficas de Usuario.	Pág (449-453)
Calidad de información en comunidades virtuales.	Pág (454-458)
Accesibilidad web. Estrategias y aplicaciones.	Pág (459-463)
Estimación del tamaño de un sistema mediante el análisis de diagramas UML y la aplicación de métodos basados en puntos de función.	Pág (464-467)
Gobernanza de Sistemas y Tecnologías de la Información (SI/TI) en el ámbito Universitario.	Pág (468-474)
Transformación Digital – Propuestas de solución para las Ciudades Inteligentes Sostenibles, la Gobernanza Digital y las Buenas prácticas y calidad.	Pág (475-480)

ISS – Innovación en Sistemas de Software	Pág (481-573)
Estado actual de la industria 4.0 en las PyMES del partido de Hurlingham.	Pág (482-486)
Sistema de recomendación ubicuo para contribuir a la disminución del estrés académico en estudiantes universitarios.	Pág (487-490)
El avance en la madurez digital de las pymes a través de la transformación digital.	Pág (491-495)
Tecnologías de la información facilitadoras para la interoperabilidad de software en Gobierno Abierto: Segunda Parte.	Pág (496-500)
Desarrollo de una app móvil de Nueva Generación.	Pág (501-504)
Soluciones IoT desde Smart Cities a Blockchain.	Pág (505-510)
Diseñar una solución de identidad auto-gestionada para acceso a servicios de calidad con redes Blockchain multipropósito en la Universidad Nacional de Río Negro.	Pág (511-515)
Aplicación de geofencing para el cuidado de adultos mayores mediante un hardware dedicado.	Pág (516-519)
Programa integrador de proyectos de Software con transferencia a la comunidad.	Pág (520-523)
Innovación en las Ciudades mediante las TIC Análisis de implementaciones en Ciudades Inteligentes.	Pág (524-528)
Aplicación de técnicas de inteligencia artificial en la prevención de problemas de salud mental.	Pág (529-533)
Soluciones de IOT: ámbitos de aplicación y desafíos.	Pág (534-538)
Machine learning y deep learning en la interpretación del lenguaje de señas.	Pág (539-543)
Diseño de una estrategia para la gestión del tiempo en la comunicación de corrección de fallas en el mantenimiento industrial.	Pág (544-548)
Implementación de una BCI mediante el diseño y desarrollo de Serious Games para la rehabilitación de pacientes neurológicos.	Pág (549-553)
La Inteligencia Artificial y su aplicación a los servicios de Internación Domiciliaria.	Pág (554-558)
Abordajes científico tecnológicos basados en el uso de vants.	Pág (559-563)
Innovación tecnológica a través de la hiper automatización.	Pág (564-568)
Tecnologías de Datos Espaciales, Visualización y Realidad Virtual.	Pág (569-573)
PDP – Procesamiento Distribuido y en Paralelo	Pág (574-636)
Paralelización de algoritmos y evaluación de rendimiento en Plataformas de cómputo de altas prestaciones.	Pág (575-580)
Análisis y modelado de sistemas utilizando Cloud Computing, IaaS privados y públicos.	Pág (581-585)
Estrategias de modelado y simulación en sistemas de HPC y salud.	Pág (586-590)
Arquitecturas Multiprocesador en HPC: Software de Base, Modelos y Aplicaciones.	Pág (591-596)
Arquitecturas Edge-Fog-Cloud en Procesamiento Distribuido. Aspectos de Eficiencia y Resiliencia.	Pág (597-602)
Soporte Serverless para Aplicaciones Móviles de Nueva Generación.	Pág (603-607)
Predicción de fenómenos naturales mediante paralelismo, algoritmos evolutivos y búsqueda por novedad.	Pág (608-612)
Simulación y análisis eficiente de grandes volúmenes de datos.	Pág (613-617)
Autoescalado basado en Aprendizaje Profundo por Refuerzo de Workflows Científicos en la Nube.	Pág (618-622)
Análisis de rendimiento de aplicaciones y sistemas complejos mediante computación paralela y técnicas de optimización de recursos.	Pág (623-627)
Técnicas de Recuperación y Procesamiento de Grandes Volúmenes de Datos Sísmicos. Un Repositorio Público Basado en Serverless.	Pág (628-631)
Aplicaciones para la salud y métricas de rendimiento en sistemas paralelos y distribuidos.	Pág (632-636)

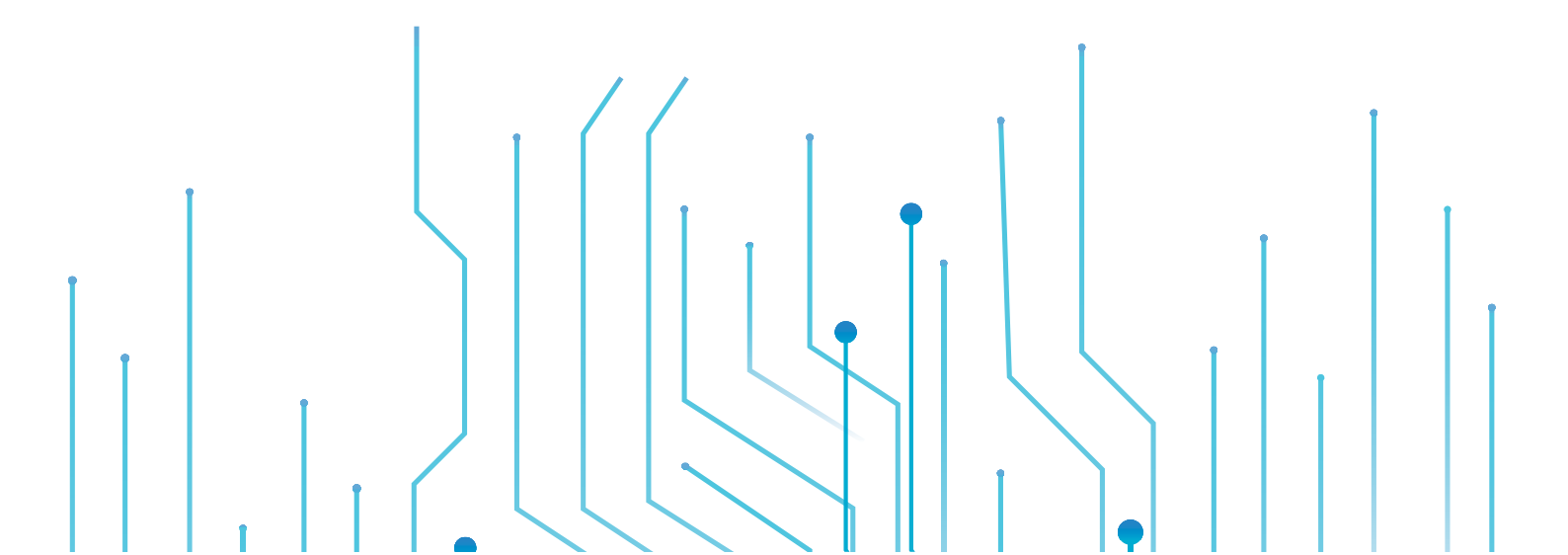
PSSTR – Procesamiento de Señales y Sistemas en Tiempo Real	Pág (637-655)
Avances en el desarrollo de un prototipo experimental de segmento terreno Satelital multiplataforma-multimisión UGS UNLaM.	Pág (638-641)
Partición de Visual SLAM para su uso con celulares.	Pág (642-646)
Implementación y Simulación de Sistemas de Tiempo Real.	Pág (647-650)
Captura de variables para alerta temprana de incendios forestales y su almacenamiento para integración a sistemas de predicción mediante utilización de redes inalámbricas de sensores.	Pág (651-655)
RCCI – Redes de Cooperación Científica Internacionales	Pág (656-666)
Maestría en Gestión y Tecnología de Ciudades Inteligentes Resultado del Proyecto CAP4CITY.	Pág (657-660)
Avances del Consorcio para la colaboración en I+D+I en Temas de Cloud Computing, Big Data y Emerging Topics (CCC-BD&ET).	Pág (661-666)
SI – Seguridad Informática	Pág (667-726)
Estudio de las propiedades criptológicas de los Tokens No Fungibles Inteligentes para asegurar dispositivos de Internet de las Cosas.	Pág (668-673)
Análisis de las Herramientas para Realizar Pruebas Estáticas de Seguridad de las Aplicaciones.	Pág (674-679)
Tecnología Blockchain Aplicada a la Ciberdefensa.	Pág (680-683)
Aprendizaje por Refuerzo Aplicado al Etiquetado de Tráfico de Red.	Pág (684-688)
Criptografía Liviana y Ciberseguridad aplicados a Sistemas Ciberfísicos.	Pág (689-693)
Software para recolección de evidencias digitales rápidas en sistemas Windows.	Pág (694-698)
Protección de activos de software.	Pág (699-703)
Técnicas para Incrementar la Seguridad en Web Services basados en WSDL.	Pág (704-708)
Avances sobre la Especificación Integral del Sistema OTP-Vote Orientada a su Implementación.	Pág (709-713)
Criptología Maliciosa para la Ciberdefensa.	Pág (714-718)
Métodos, Técnicas y Herramientas para la Protección de Sistemas de Software.	Pág (719-723)
La blockchain y la certificación de documentos en Argentina.	Pág (724-726)
TD – Tesis Doctorales	Pág (727-775)
Generalización del modelado de cadencias de tecleo con contextos finitos para su utilización en ataques de presentación y canal lateral.	Pág (728-737)
Cálculo Científico Distribuido sobre Clientes Móviles Indeterminados en redes MANETs.	Pág (738-747)
DepProMod: Modelo de Proceso de Despliegue de Sistemas de Software.	Pág (748-757)
Tesis Doctoral. Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales.	Pág (758-767)
Modelo de madurez para servicios de gobierno electrónico en el ámbito universitario.	Pág (768-775)
TIAE – Tecnología Informática Aplicada en Educación	Pág (776-897)
Tecno pedagogía en el Proceso Formativo.	Pág (777-781)
Estrategias de accesibilidad para una educación inclusiva en la UNSAdA.	Pág (782-786)
Diseño de herramientas tecnológicas en Modelos educativos híbridos.	Pág (787-791)
Recursos educativos accesibles. Estudio de distintas técnicas y herramientas basadas en inteligencia artificial.	Pág (792-797)
Experiencia de Usuario en Sistemas Interactivos usando reconocimiento de Emociones.	Pág (798-802)

Un bot virtual en Moodle para asistir y complementar actividades académicas en espacios de educación superior.	Pág (803-807)
Interacción Persona Ordenador y Educación. Desarrollos y experiencias.	Pág (808-813)
Realidad aumentada en contextos educativos y su relación con el rendimiento académico universitario.	Pág (814-815)
Tecnologías digitales aplicadas al escenario educativo. Desarrollos y experiencias.	Pág (816-820)
Tecnologías disruptivas y aulas emergentes en el nivel medio y superior en el contexto de la UNRN.	Pág (821-825)
Desarrollo y aplicación de laboratorios virtuales destinados a educación de grado y pregrado.	Pág (826-830)
La mejora del aprendizaje en AVAs aplicando analítica del aprendizaje.	Pág (831-835)
La incorporación de la Ciencia de Datos como estrategia transversal para reforzar la enseñanza de la programación en el aula.	Pág (836-840)
Realidad Virtual y Educación.	Pág (841-845)
Diagnóstico de competencias genéricas universitarias, a través de software business intelligence.	Pág (846-850)
Revisión de alternativas moodle para la extracción de indicadores de autorregulación del aprendizaje.	Pág (851-855)
Realidad virtual para la enseñanza de portugués: Una guía para la implementación de contenido educativo en estudiantes de nivel medio.	Pág (856-860)
Definición de Lenguajes Específicos del Dominio para la Generación de Sitios Web de Instituciones Escolares.	Pág (861-865)
Definición de Lenguajes Específicos del Dominio para la Generación de Sitios Web de Instituciones Escolares.	Pág (866-869)
Minisatélite CANSAT para la enseñanza de las ciencias experimentales en diferentes contextos educativos.	Pág (870-874)
Calidad educativa: instrumentos para alcanzar la mejora continua de la educación superior.	Pág (875-879)
Evaluación del desempeño docente en el nivel superior.	Pág (880-884)
Software para gamificar terapias de rehabilitación AD-HOC.	Pág (885-890)
Tecnologías innovadoras aplicadas en contextos educativos	Pág (891-894)
Alfabetizando: innovación en el proceso de enseñanza aprendizaje en escuelas secundarias usando visualización de datos para desarrollar pensamiento computacional en materias de Humanidades y Ciencias Sociales.	Pág (895-897)
Autores:	Pág (898-901)



BDDM

Bases de Datos y Minería de Datos



Propuesta de técnicas de validación para la calidad de datos abiertos e identificación de patrones para predicciones con Machine Learning:

Segunda Parte

Roxana Martínez, Pablo Vilaboa, Agustín Simón, Juan Iannini, Eugenio Clavijo, German Colato, Nelson Catala, Carlos Bateca, Serenela Lopumo, Hernán Zabala, Gonzalo Parada, Ayelén Chaglasian Sgang, Rocío Vargas, Sebastián Tsiro

CAETI - Centro de Altos Estudios en Tecnología Informática
Universidad Abierta Interamericana (UAI)
Montes de Oca 745, Ciudad Autónoma de Buenos Aires, Argentina

{roxana.martinez, pablo.vilaboa, rocio.vargas}@uai.edu.ar

{MiguelAgustin.Simon, Juanmanuel.Garciaianini1, Eugenio.Clavijo, GermanDario.Colato, Nelson.Catala, CarlosAlberto.BatecaCaicedo, Serenela.Lopumo, HernanGuillermo.Zabala, Gonzalo.Parada, AyelenMalena.Chaglasian, SebastianMartinDemetrio.Tsiros}@alumnos.uai.edu.ar

RESUMEN

Dentro de los varios focos del paradigma de Gobierno Abierto, existen los conjuntos de datos abiertos públicos disponibilizados en los sitios web gubernamentales, éstos permiten no sólo una mejor transparencia en las actividades públicas que se realizan, sino que, además, pueden ser reutilizados a través de estudios para brindar mejores servicios y obtener beneficios para la sociedad. Mantener estos datos con calidad, permite que se lleven a cabo, análisis de escenarios para predecir y evitar posibles falencias o errores en la toma de decisiones políticas que tenga un alto impacto en los ciudadanos. Actualmente, no existen demasiados trabajos de investigación que se centren en el contenido de los datasets, en aspectos de: análisis de contenido interno, validación, herramientas de verificación, cotejo en las estructuras de datos, guías de buenas prácticas, o bien reconocimiento de patrones estructurales, entre otros temas. Es por esto, por lo que, este trabajo de investigación se centra en dichos puntos mencionados para validar la calidad de los datos públicos abiertos y que éstos puedan ser utilizados con un fin positivo en la comunidad. Este trabajo, no solo trabaja en prototipos de software en validaciones de

datos específicos, sino que, además, permite estudiar modelos predictivos, que a través del aprendizaje automático pueden brindar un mejor análisis para la toma de decisiones.

Palabras clave: Datos Abiertos, Gobierno Abierto, Métricas de Calidad de Datos, Machine Learning, Predicciones a partir de Patrones.

CONTEXTO

El presente trabajo es la segunda parte del proyecto denominado “Investigación y desarrollo de software para la validación de la calidad de datos abiertos e identificación de patrones para predicciones”, que tiene inicio en el mes de marzo 2022 y con fecha de culminación en el mes de marzo 2024. Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI) de la Facultad de Tecnología Informática de la Universidad Abierta Interamericana (UAI), el cual contribuye al desarrollo de las Tecnologías de la Información y las Comunicaciones (TIC) en Argentina y en el mundo, llevando adelante la investigación en diversas áreas a través de tratamiento y manipulación de datos en diversas temáticas, tales como: aspectos gubernamentales,

gobierno de datos, economía, salud, videojuegos, multimedia entre otras. El proyecto es financiado y evaluado por la Secretaría de Investigación de la Universidad, tiene una duración de 2 años, y cuenta con la participación de docentes y estudiantes de grado y posgrado en diversas carreras de la Facultad de Tecnología Informática.

1. INTRODUCCIÓN

Gobierno Abierto se encuentra en pleno auge tanto a nivel nacional como internacional, haciendo partícipe a distintos gobiernos y ciudadanos en este contexto colaborativo y participativo de datos y servicios, en conjunto con el objetivo de brindar una mayor transparencia en las acciones gubernamentales realizadas por parte del Estado Nacional. Si bien Gobierno Abierto implica varios enfoques, este trabajo se centra en la implementación de nuevas técnicas de gestión de datos y en el estudio de la disponibilización de datos públicos a través de un diseño tecnológico, que se encuentra centrado en el ciudadano promedio para fomentar la inclusión social. “Al ampliar el acceso a la información pública se fortalece la rendición de cuentas y se enriquece el debate público, a la vez que se crean nuevas oportunidades para generar valor agregado” [1]. Por otra parte, es necesario mencionar que “los datos que se pueden reutilizar y redistribuir sin ninguna restricción se denominan datos abiertos” [2]. Estos se encuentran en los portales gubernamentales o bien en sitios web que poseen licencias específicas que permiten trabajar con formatos digitales concretos, sobre un posible modelo estándar abierto.

Es preciso conocer que “la cultura de datos abiertos trata de obtener un valor añadido de la información [...], la información no genera sólo su valor por estar reservada a unos pocos, sino que lo hace por su disponibilidad para ser interpretada y traducida por cualquier actor interesado en trabajar con ella” [3], es decir, un dato abierto puede ser accedido a través de un formato especial que permite la interoperabilidad con otros softwares.

Como parte de la problemática de este tema, cabe mencionar que “no toda la información pública disponible o publicada en la web es información abierta válida para su reutilización. No sólo se trata de publicar los datos, sino que hay que garantizar el acceso a ellos, razón por la que debe recurrirse a formatos digitales, estandarizados y abiertos, siguiendo una estructura clara que permita su comprensión y reutilización” [4].

Existen algunos trabajos enfocados a las mediciones en aspectos de calidad de datos públicos [5] [6] [7], por otro lado, otras investigaciones apuntan a medidas otorgadas a la calidad de los datos abiertos, este es el caso del esquema modelo de cinco estrellas de Berners-Lee [8], o bien trabajos realizados con informes sobre el estado de los datos elaborados por fundaciones comprometidas con el impacto de estos datos públicos [9]. Es importante que se trabaje de manera colaborativa sobre algunos puntos FODA (Fortalezas, Oportunidades, Debilidades y Amenazas) con respecto a varias situaciones a nivel Estado Nacional, con el fin que tanto las entidades gubernamentales como los ciudadanos, puedan brindar soluciones específicas a determinados problemas trabajando en conjunto [10].

Gracias al gran impulso de la tecnología en el área de datos, el concepto de Machine Learning (ML) cada vez se encuentra más en auge, también conocido en español como aprendizaje automático, es una línea que deriva de la inteligencia artificial que hace foco en estudiar y crear algoritmos que permiten entrenar los procesos de decisión de las computadoras. En el contexto del aprendizaje automático, los datos son un insumo importante, ya que son el punto de inicio desde el cual se puede aprender y entrenar cada uno de los modelos de Machine Learning [11]. En lo que respecta al proceso de entrenamiento, implica linealmente el uso de datos, los cuales pueden ser datos estructurados o no estructurados [...]. La elección de usar un algoritmo depende del tipo de problema que se esté abordando. Esta decisión, es parte del proceso de selección del ingeniero de datos, que tiene como tarea,

identificar y trabajar con el algoritmo que mejor se adapte a cada problema [12].

El aprendizaje automatizado, tiene una amplia variedad de usos en diferentes ámbitos. Algunos de los beneficios de su utilización, se muestran a continuación:

- En el ámbito Financiero, ML se usa para la detección de fraudes, la evasión, el análisis de perfil para otorgar créditos, entre otros aspectos. Los modelos, analizan grandes volúmenes de datos de fuentes diferentes. Por ejemplo, los algoritmos cruzan los datos financieros buscando patrones o anomalías que pudiesen indicar una o varias actividades de interés para la investigación [13].
- Otro ámbito es el contexto de la salud: el aprendizaje automático es usado en varios puntos, por ejemplo: puede utilizarse para predecir enfermedades y analizar imágenes médicas. Del mismo modo, estos modelos analizan grandes volúmenes de datos médicos para interpretar el comportamiento de algunas variables en la búsqueda de nuevo conocimiento con mayor rapidez y precisión [14].
- El sector del marketing utiliza machine learning para personalizar anuncios y para establecer una segmentación adecuada del público objetivo. Los modelos de aprendizaje automatizado pueden analizar los datos de cada cliente identificando comportamientos genéricos y preferencias según su segmentación. Esto permite que los profesionales puedan crear anuncios dirigidos a un cliente potencial y con ello construir un canal de comunicación más efectivo [15] [16].
- El sector de la producción utiliza aprendizaje automatizado para diferentes tareas dentro de la empresa como el mantenimiento predictivo, la optimización de los procesos de producción, entre otros. Los modelos de ML analizan datos de los sensores de la planta de producción, los datos se recopilan durante el proceso y se almacenan. Con estos datos, se pueden predecir fallos, o se puede mejorar el proceso para hacer un mejor uso de los recursos. El efecto de utilizar proceso de

aprendizaje automatizado permite reducir el tiempo de inactividad y los costos que estas fallas producen al realizar tareas de mantenimiento predictivo [17].

En conclusión, la utilización del aprendizaje automatizado permite trabajar con los datos para entrenar a diferentes modelos, que luego, éstos a través de algoritmos, colaboran en la toma de decisiones en el estudio de escenarios de análisis. Su aplicación ha conquistado muchos ámbitos, ofreciendo beneficios y usos variados. Esto último, hace de este tema un campo de estudio muy interesante. Por otra parte, es fundamental comprender que para que esto se consiga, es necesario contar con datos de calidad en los portales de datos abiertos que serán utilizados en contextos de ML. Sin esta calidad, los resultados estudiados podrían verse sesgado o con falencias. Hoy por hoy, sigue siendo un gran desafío para las políticas públicas, ya que muchos de los conjuntos de datos públicos no se encuentran en formatos estandarizados o bien se presentan incompletos.

2. LÍNEAS DE INVESTIGACIÓN DESARROLLO

Este proyecto pertenece a la línea de investigación de Ingeniería de Software del Centro de Altos Estudios en Tecnología Informática (CAETI). Para esta segunda parte del proyecto, se hace enfoque en los siguientes ejes del tema:

- Continuar con el diseño y desarrollo de algoritmos para identificar patrones en datos públicos abiertos.
- Estudiar modelos predictivos orientados a contextos gubernamentales, como así también, analizar posibles predicciones sobre nuevos datos encontrados con técnicas de machine learning.
- Analizar herramientas de manejo con Machine Learning que permitan análisis de patrones con enfoque predictivo.
- Establecer criterios para mantener estándares de calidad de datos;
- Desarrollar una herramienta de validación de datasets gubernamentales basada en

métricas de calidad de datos estándares como geoespaciales;

- Generar una guía de buenas prácticas para las técnicas de Machine Learning en la utilización de datos abiertos.
- Definir y recolectar la muestra de datasets para ser testeados por la herramienta propuesta como validadora de calidad;
- Análisis de casos predictivos con casos reales.
- Análisis de casos de aplicación con los prototipos desarrollados.

3. RESULTADOS OBTENIDOS/ESPERADOS

Enfoque del proyecto:

Los conjuntos de datos públicos del contexto ofrecido por el nuevo paradigma de Gobierno Abierto, si bien son una fuente útil que se disponibiliza a los distintos organismos, como ciudadanos, tienen falencias en aspectos de calidad de los datos. Es por lo que resulta necesario, que éstas puedan ser detectadas y analizadas a tiempo, para evitar errores en el estudio de éstos, y mitigar posibles interpretaciones erróneas. Para ello, la utilización de herramientas de software para realizar la verificación y medición de este aspecto son fundamentales. Mantener la calidad de datos, permitirá trabajar con algoritmos en los que se pueda detectar patrones que fomenten el estudio de predicciones en escenarios posibles, y que esto traerá aparejado, beneficios sociales a las personas.

Objetivos principales:

Realizar el análisis, diseño y desarrollo de herramientas de software para la gestión y validación de la calidad de los datos públicos en el contexto de Gobierno Abierto. Detectando el “estado de salud” de las diversas fuentes de datos provenientes de casos de aplicación gubernamentales con los prototipos desarrollados, incorporando algoritmos para identificar patrones que logren predicciones sobre nuevos datos.

Metodología y Técnicas:

El proceso metodológico de investigación utilizado para esta segunda parte del proyecto, continúa con un proceso sistemático cualitativo, que implica las siguientes etapas: estudio del problema; revisión teórica; recolección y revisión de datos, clasificación y análisis de datos; estudio de escenarios de identificación de patrones sobre éstos con enfoques predictivos sobre nuevos datos; desarrollo de prototipos (análisis cuantitativo de datos); visualización de la predicción de patrones; validación de la solución propuesta; identificación de las limitaciones del trabajo.

Resultados obtenidos:

En lo que fue la primera parte del proyecto (primer año), se analizaron las falencias en aspectos de calidad de datos públicos abiertos, se lograron identificar técnicas de validación de calidad de datos abiertos en formatos estándares y se llevó a cabo el estudio de datos geolocalizados que se encuentran disponibilizados en los portales gubernamentales, como así también, se trabajó en el análisis de propuestas de diseño y desarrollo de algoritmos con la finalidad de descubrir patrones en datos públicos. Además, se estudiaron posibles modelos predictivos orientados a este contexto (herramientas de software y demás), para efectuar predicciones y mejorar los resultados de análisis en resultados finales con técnicas de Machine Learning. Finalmente, se lanzó un primer prototipo de software, que tiene como objetivo la validación de calidad en datos públicos abiertos geolocalizados.

4. FORMACIÓN DE RECURSOS HUMANOS

Este proyecto se compone por una docente investigadora, Doctora en Ciencias Informáticas de la Universidad Nacional de La Plata (UNLP), directora del Proyecto de Investigación, 2 (dos) docentes tesistas y 2 (dos) estudiantes tesistas de la carrera de Maestría en Tecnología Informática en la Universidad Abierta Interamericana (UAI). También cuenta con un docente tesista de la

carrera del Doctorado en Ciencias Informáticas de UAI, un estudiante tesista de la carrera de la Especialización en Ingeniería de Softwares de UAI, 5 (cinco) estudiantes colaboradores de la carrera de Ingeniería en Sistemas Informáticos de UAI, y 2 (dos) tesistas de la carrera Licenciatura en Gestión de Tecnología Informática en UAI. En relación directa con la línea de I/D presentada para el proyecto, los miembros del equipo se encuentran en realización de: 1 tesis doctoral, 4 tesis de maestría, 1 tesis de especialización y 2 tesinas de grado de la UAI.

5. BIBLIOGRAFÍA

- [1] Buenos Aires provincia (2017). “Kit de Apertura Municipal”. Disponible en: <http://escueladefiscales.com/Kit%20de%20Apertura%20Municipal%202017%20-%20provincia%20de%20buenos%20aires.pdf>
- [2] Oviedo, E., Mazón, J. N., & Zubcoff, J. J. (2013). Hacia un modelo de calidad de datos para portales de datos abiertos. In XXXIX Latin American Computing Conference (CLEI), Nanguata (pp. 1-8).
- [3] Magallón Rosa, R. (2017). Datos abiertos y acceso a la información pública en la reconstrucción de la historia digital.
- [4] Garriga-Portolà, M. (2011). ¿Datos abiertos? Sí, pero de forma sostenible. Profesional de la Información, 20(3), 298-303.
- [5] ISO 25012 (2008). “Ingeniería de software - Requisitos de calidad y evaluación de productos de software (SQuaRE) - Modelo de calidad de datos”. Disponible en: <https://www.iso.org/obp/ui/es/#iso:std:iso-iec:25012:ed-1:v1:en>
- [6] Martínez, R. et al. (2021). Metrics proposal to measure the quality of governmental datasets. IEEE Latin America Transactions, Vol. 100. ISSN 1548-0992.
- [7] de España, G. (2017). Manual práctico para mejorar la calidad de los datos abiertos. Madrid. Disponible en: https://datos.gob.es/sites/default/files/doc/file/manual_practico_para_mejorar_la_calidad_de_los_datos_abiertos_1.pdf
- [8] Open Data (2012). “5 Open Data”. Disponible en: <https://5stardata.info/en/>
- [9] ODI Open Data Institute. “The 2019 Data Skills Framework”. Disponible en: <https://theodi.org/article/open-data-skills-framework/>
- [10] Martínez, R. (2022). “Métricas de calidad para validar los conjuntos de datos abiertos públicos gubernamentales” (Doctoral dissertation, Universidad Nacional de La Plata).
- [11] Baloglu, O., Latifi, S. Q., & Nazha, A. (2022). “What is machine learning?”. Archives of Disease in Childhood-Education and Practice, 107(5), 386-388.
- [12] Mahesh, B. (2020). “Machine Learning Algorithms - A Review”. International Journal of Science and Research (IJSR), 9, 7.426. doi:10.21275/ART20203995.
- [13] Caneiro, L. P., & Ortega, E. R. (2021). “Predicción de series temporales basada en Machine Learning: aplicaciones económicas y financieras. In Nuevos métodos de predicción económica con datos masivos” (pp. 189-214). Fundación de las Cajas de Ahorros (FUNCAS).
- [14] Oala, L., Murchison, A. G., Balachandran, P., Choudhary, S., Fehr, J., Leite, A. W., ... & Wiegand, T. (2021). “Machine learning for health: algorithm auditing & quality control”. Journal of medical systems, 45, 1-8.
- [15] Brei, V. A. (2020). “Machine learning in marketing: Overview, learning strategies, applications, and future developments”. Foundations and Trends® in Marketing, 14(3), 173-236.
- [16] Mithun S, U., Thonse, H., Rashmi, S., & hamed, N. (2021). “El papel del aprendizaje automático en el marketing digital”. SAGE Open.
- [17] Weichert, D., Link, P., Stoll, A., Rüping, S., Ihlenfeldt, S., & Wrobel, S. (2019). “A review of machine learning for the optimization of production processes”. The International Journal of Advanced Manufacturing Technology, 104(5-8), 1889-1902.