



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Departamento:
Departamento de Ingeniería e Investigaciones Tecnológicas

Programa de acreditación:
PROINCE

Programa de Investigación¹:

Código del Proyecto: C230

Título del proyecto
Aprendizaje automático para el control del estado de salud en sistemas Aeroespaciales

PIDC:

Elija un elemento.

PII:

Elija un elemento.

Director: Ierache, Jorge Salvador

Codirector: Soligo, Pablo

Integrantes Alumnos:
Merkel, German

Investigador Externo, Asesor- Especialista, Graduado UNLaM

Resolución Rectoral de acreditación: N° 260/20

Fecha de inicio: 01/01/2020

Fecha de finalización: 31/12/2021

¹ Los Programas de Investigación de la UNLaM están acreditados con resolución rectoral, según lo indica la Resolución HCS N° 014/15 sobre **Lineamientos generales para el establecimiento, desarrollo y gestión de Programas de Investigación a desarrollarse en la Universidad Nacional de La Matanza**. Consultar en el departamento académico correspondiente la inscripción del proyecto en un Programa acreditado.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

A. Desarrollo del proyecto (adjuntar el protocolo)

A.1. Grado de ejecución de los objetivos inicialmente planteados, modificaciones o ampliaciones u obstáculos encontrados para su realización (desarrolle en no más de dos (2) páginas)

Durante el primer año de proyecto y debido a las restricciones para la participación en congresos se priorizaron los puntos del plan de proyecto vinculados a la publicación del UGS (<https://ugs.unlam.edu.ar>) (Unlam Ground Segment) y al sitio del grupo de investigación (<https://gidsa.unlam.edu.ar/>). Los objetivos han sido mayormente cumplidos durante los dos años del proyecto. Los resultados parciales han sido publicados en los congresos y workshops detallados en el presente documento, (Soligo, Ierache, & Merkel, Telemetría de altas prestaciones sobre base de datos de serie de tiempos, 2020), (Soligo, Merkel, & Ierache, Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital, 2021) y (Soligo, Ierache, & Merkel, Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia., 2021).

El siguiente cuadro muestra los objetivos específicos planteados y su grado de ejecución.

Objetivo	Cumplimiento		
	Total	Parcial	No cumplido
Identificar correlaciones ocultas entre los datos	X		
Proponer los mejores modelos según el caso y el rendimiento observado	X		
Evaluar y, de ser posible implementar, automatización en la selección de modelos		X	
Lograr una solución práctica, autónoma y libre de mantenimiento.		X	
Formalizar el proceso y generar resultados medibles que justifiquen su implementación	X		
Evaluar tecnologías de persistencia alternativas	X		
Prueba piloto estación terrena UNLaM	X		
Publicación del UGS	X		

El trabajo (Soligo, Ierache, & Merkel, Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia., 2021) demuestra el potencial del aprendizaje automático para la detección temprana de anomalías. En muchos casos estas no pueden ser detectadas con un simple control de límites y el aprendizaje automático, sensible al contexto, los estados y las consecuentes correlaciones pueden ser un aliado de valor. La nula disponibilidad de conjuntos de datos etiquetados y la ausencia de datos históricos en misiones de nanosatélites ha sido el principal obstáculo a la hora de profundizar en estas técnicas. La selección de modelos ha sido un proceso manual, aunque si se pudo automatizar la parametrización sobre un modelo en particular.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Las ampliaciones más destacadas respecto a la planificación original se detallan a continuación:

Tecnologías de persistencia alternativas

La cantidad de datos acumulados durante la publicación del prototipo UGS(Unlam Ground Segment) (<https://ugs.unlam.edu.ar>) y su funcionamiento 7x24 potenciaron la búsqueda de alternativas de bajo costo para el almacenamiento y explotación de telemetría como medida preventiva al aumento de tamaño de tablas de almacenamiento de telemetría. Los resultados de la investigación están disponibles en el artículo (Soligo, Ierache, & Merkel, Telemetría de altas prestaciones sobre base de datos de serie de tiempos, 2020).

Prueba piloto estación terrena UNLaM

Se obtuvo de forma directa y durante una pasada el primer paquete del satélite argentino Bussat1/TITA por medio de equipamiento y tecnología propia. Mediante una antena de origen nacional y con equipamiento previamente adquirido se obtuvo un paquete/baliza del mencionado satélite. El paquete fue demodulado, decodificado e interpretado correctamente. Actualmente se puede visualizar en el UGS.

<https://ugs.unlam.edu.ar/#/browse/mine/demodash?view=tabs&tc.mode=fixed&tc.startBound=1621281857194.7888&tc.endBound=1621282757194.7888&tc.timeSystem=utc&tabs.pos.demodash=0>

Este hito representa el primer paso para que la UNLaM tenga su propia estación terrena y domine todo el ciclo de ingeniería.

Publicación del UGS

Se instaló en servidores de la UNLaM el UGS. Ya se encuentra público y accesible en <https://ugs.unlam.edu.ar>

B. Principales resultados de la investigación

B.1. Publicaciones en revistas (informar cada producción por separado)

Artículo 1:	
Autores	<i>Pablo Soligo, Jorge Salvador Ierache, Germán Merkel</i>
Título del artículo	<i>Telemetría de altas prestaciones sobre bases de datos de serie de tiempos.</i>
N° de fascículo	2
N° de Volumen	5
Revista	<i>Revista Digital del Departamento de Ingeniería e Investigaciones Tecnológicas de la UNLaM.</i>



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Año	2020
Institución editora de la revista	<i>Facultad de Ciencias Exactas, Físicas y Naturales Universidad Nacional de San Juan</i>
País de procedencia de institución editora	<i>Argentina</i>
Arbitraje	SI
ISBN:	2525-1333
URL de descarga del artículo	https://reddi.unlam.edu.ar/index.php/ReDDi/article/view/131/252
N° DOI	-

Artículo 2:	
Autores	<i>Pablo Soligo, Jorge Salvador Ierache, Germán Merkel</i>
Título del artículo	<i>Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital</i>
N° de fascículo	-
N° de Volumen	-
Revista	<i>XXIII Workshop de Investigadores en Ciencias de la Computación</i>
Año	2021
Institución editora de la revista	<i>Red de Universidades con Carreras en Informática</i>
País de procedencia de institución editora	<i>Argentina</i>
Arbitraje	SI
ISBN:	978-987-24611-3-3; 978-987-24611-4-0
URL de descarga del artículo	http://se-dici.unlp.edu.ar/handle/10915/120207
N° DOI	

Artículo 3:	
Autores	<i>Pablo Soligo, Jorge Salvador Ierache, Germán Merkel</i>
Título del artículo	<i>Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia.</i>
N° de fascículo	-
N° de Volumen	-



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Revista	<i>Libro de Actas XXVII Congreso Argentino de Ciencias de la Computación CACIC 2021</i>
Año	2021
Institución editora de la revista	<i>Universidad Nacional de Salta</i>
País de procedencia de institución editora	<i>Argentina</i>
Arbitraje	SI
ISBN:	978-987-633-574-4
URL de descarga del artículo	http://se-dici.unlp.edu.ar/bitstream/handle/10915/130454/Documento_completo.pdf?sequence=1&isAllowed=y
N° DOI	-

B.2. Libros

Libro 1	
Autores	
Título del Libro	
Año	
Editorial	
Lugar de impresión	
Arbitraje	Elija un elemento.
ISBN:	
URL de descarga del libro	
N° DOI	

B.3. Capítulos de libros

Autores	
Título del Capítulo	
Título del Libro	
Año	
Editores del libro/Compiladores	
Lugar de impresión	
Arbitraje	Elija un elemento.
ISBN:	
URL de descarga del capítulo	



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

N° DOI	
--------	--

B.4. Trabajos presentados a congresos y/o seminarios

Autores	<i>Pablo Soligo, Jorge Salvador Ierache, Germán Merkel</i>
Título	<i>Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital</i>
Año	2021
Evento	<i>XXIII Workshop de Investigadores en Ciencias de la Computación</i>
Lugar de realización	<i>Universidad Nacional de Chilecito</i>
Fecha de presentación de la ponencia	15/04/2021
Entidad que organiza	<i>Universidad Nacional de Chilecito (Virtual)</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	http://se-dici.unlp.edu.ar/handle/10915/120207

Autores	<i>Pablo Soligo, Jorge Salvador Ierache, Germán Merkel</i>
Título	<i>Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia.</i>
Año	2020
Evento	<i>Congreso Argentino de Ciencias de la Computación CACIC 2021</i>
Lugar de realización	<i>Universidad Nacional de Salta (Virtual)</i>
Fecha de presentación de la ponencia	01/10/2021
Entidad que organiza	<i>Universidad Nacional de Salta</i>
URL de descarga del trabajo (especificar solo si es la descarga del trabajo; formatos pdf, e-pub, etc.)	http://se-dici.unlp.edu.ar/bitstream/handle/10915/130454/Documento_completo.pdf?sequence=1&isAllowed=y



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

B.5. Otras publicaciones

Autores	
Año	
Título	
Medio de Publicación	

C. Otros resultados. Indicar aquellos resultados pasibles de ser protegidos a través de instrumentos de propiedad intelectual, como patentes, derechos de autor, derechos de obtentor, etc. y desarrollos que no pueden ser protegidos por instrumentos de propiedad intelectual, como las tecnologías organizacionales y otros. Complete un cuadro por cada uno de estos dos tipos de productos.

C.1. Títulos de propiedad intelectual. Indicar: Tipo (marcas, patentes, modelos y diseños, la transferencia tecnológica) de desarrollo o producto, Titular, Fecha de solicitud, Fecha de otorgamiento

Tipo	Titular	Fecha de Solicitud	Fecha de Emisión

C.2. Otros desarrollos no pasibles de ser protegidos por títulos de propiedad intelectual. Indicar: Producto y Descripción.

Producto	Descripción

D. Formación de recursos humanos. Trabajos finales de graduación, tesis de grado y posgrado. Completar un cuadro por cada uno de los trabajos generados en el marco del proyecto.

D.1. Tesis de grado

Director (apellido y nombre)	y Autor (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.2 Trabajo Final de Especialización

Director (apellido y nombre)	y Autor (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título del Trabajo Final

D.2. Tesis de posgrado: Maestría



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Director (apellido y nombre)	Tesista (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.3. Tesis de posgrado: Doctorado

Director (apellido y nombre)	Tesista (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título de la tesis

D.4. Trabajos de Posdoctorado

Director (apellido y nombre)	Posdoctorando (apellido y nombre)	Institución	Calificación	Fecha /En curso	Título del trabajo	Publicación

E. Otros recursos humanos en formación: estudiantes/ investigadores (grado/posgrado/ posdoctorado)

Apellido y nombre del Recurso Humano	Tipo	Institución	Período (desde/hasta)	Actividad asignada ²
German Merkel	Estudiante	UNLaM	01/01/2020-31/12/2021	Becario investigador. Investigación sobre fuentes de telemetría satelital, simuladores, aprendizaje automático. Sincronización periódica de Nasa OPEN-MCT - UGS

Satnogs network

El UGS procesa información cruda principalmente de la comunidad satnogs (<https://satnogs.org/>).

El UGS utiliza los servicios de publicación de telemetría satelital cruda de la comunidad, decodifica la misma y la publica en <https://ugs.unlam.edu.ar>.

La comunidad ha sido informada sobre el proyecto y, si bien los datos son de libre acceso, según recomendación, se han dispuesto en varios lugares elementos que informen el origen de los datos.

² Descripción de la/s actividad/es a cargo (máximo 30 palabras)



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Se encuentra activa la búsqueda de misiones satelitales universitarias que se quieran apoyar en a la infraestructura de segmento terreno creada.

Congreso Argentino de Ciencias de la Computación

Durante el XXVI Congreso Argentino de Ciencias de la Computación realizado en la UNLaM, se ha impartido, por medio de la escuela de informática, el curso “Python de Cero a Django”. El material, las prácticas y los datos son derivados de las investigaciones del grupo GIDSA.

Semana de la Ciencia

Presentación del grupo de investigación en semana de la ciencia.

Tecnópolis

Presentación de carácter divulgativo en Tecnópolis. Se preparo el material y los contenidos, finalmente la charla fue cancelada por motivos ajenos al grupo de investigación.

G. Otra información. Incluir toda otra información que se considere pertinente.

El objetivo de mejorar la calidad de los trabajos presentados durante el 2020, Pablo Soligo, codirector del proyecto concluyo exitosamente el Curso de Escritura Científica I, organizado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de La Matanza, con una duración total de 30 horas.

Referencias

- Soligo, P., Ierache, J. S., & Merkel, G. (2020). Telemetría de altas prestaciones sobre base de datos de serie de tiempos.
- Soligo, P., Ierache, J. S., & Merkel, G. (2021). Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia. *XXVII Congreso Argentino de Ciencias de la Computación (CACIC)(Universidad Nacional de Salta, Salta. 4 al 8 de octubre de 2021)*.
- Soligo, P., Merkel, G., & Ierache, J. S. (2021). Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital. *XXIII Workshop de Investigadores en Ciencias de la Computación (WICC 2021, Chilecito, La Rioja)*.



Dr Jorge Ierache.
San Justo 22/03/2020



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

H. Cuerpo de anexos:

Anexo I

Certificados y trabajos

Curso Escritura Científica I

Curso de Escritura Científica I

Se certifica que

Sotigo, Pablo
DNI: 25.127.140

aprobó el *Curso de Escritura Científica I*, organizado por la Secretaría de Ciencia y Tecnología de la Universidad Nacional de La Matanza, con una duración total de 30 horas.

San Justo, jueves 30 de julio de 2020.


Mg. Ana Bidiña
Secretaría de Ciencia y Tecnología
UNLaM



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Dictado Curso “Python de cero a Django”, Congreso Argentina de Ciencias de la Computación 2020



REDDI 2020





Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

WICC2021



Por medio del presente se certifica que:

SOLIGO, Pablo

ha participado en calidad de Autor del artículo Nro. 13526 "Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital" aceptado en el XXIII WORKSHOP DE INVESTIGADORES EN CIENCIAS DE LA COMPUTACIÓN – WICC 2021, organizado en modalidad virtual por la Universidad Nacional de Chilecito los días 15 y 16 de abril de 2021.

Chilecito, La Rioja, Argentina.



Lic. Patricia Pesado
Coordinadora
Red UNCI



Ing. Norberto Raúl Caminoa
Rector
UNIVERSIDAD NACIONAL de CHILECITO



Por medio del presente se certifica que:

MERKEL, German

ha participado en calidad de Autor del artículo Nro. 13526 "Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital" aceptado en el XXIII WORKSHOP DE INVESTIGADORES EN CIENCIAS DE LA COMPUTACIÓN – WICC 2021, organizado en modalidad virtual por la Universidad Nacional de Chilecito los días 15 y 16 de abril de 2021.

Chilecito, La Rioja, Argentina.



Lic. Patricia Pesado
Coordinadora
Red UNCI



Ing. Norberto Raúl Caminoa
Rector
UNIVERSIDAD NACIONAL de CHILECITO



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Por medio del presente se certifica que:

IERACHE, Jorge

ha participado en calidad de Autor del artículo Nro. 13526 "Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital" aceptado en el XXIII WORKSHOP DE INVESTIGADORES EN CIENCIAS DE LA COMPUTACIÓN – WICC 2021, organizado en modalidad virtual por la Universidad Nacional de Chilecito los días 15 y 16 de abril de 2021.

Chilecito, La Rioja, Argentina.



Lic. Patricia Pesado
Coordinadora
Red UNCI



Ing. Norberto Raúl Caminoa
Rector
UNIVERSIDAD NACIONAL de CHILECITO

CACIC 2021

CACIC
2021



XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN

Por medio del presente se CERTIFICA que:

Pablo Soligo

Ha participado en calidad de AUTOR del trabajo "Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia." en el XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN llevado a cabo de manera virtual por la Facultad de Ciencias Exactas de la UNSa del 04 al 08 de octubre de 2021.

Salta, Argentina



Lic. Patricia Pesado
Coordinadora
Red UNCI



Ing. Daniel Hoyos
Decano
Facultad de Ciencias Exactas
UNSA

El presente documento fue elaborado por:
Lic. Patricia Pesado
Coordinadora
Red UNCI



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN

Por medio del presente se CERTIFICA que:

German Merkel

Ha participado en calidad de AUTOR del trabajo "Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia." en el XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN llevado a cabo de manera virtual por la Facultad de Ciencias Exactas de la UNSa del 04 al 08 de octubre de 2021.

Salta, Argentina


Lic. Patricia Pesado
Coordinadora
Red UNCI


Ing. Daniel Hoyos
Decano
Facultad de Ciencias Exactas
UNSa

Formado digitalmente por: Gil, Susana Liliana
Lic. Decano
Facultad de Ciencias Exactas
Universidad Nacional de Salta



XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN

Por medio del presente se CERTIFICA que:

Jorge Ierache

Ha participado en calidad de AUTOR del trabajo "Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia." en el XXVII CONGRESO ARGENTINO DE CIENCIAS DE LA COMPUTACIÓN llevado a cabo de manera virtual por la Facultad de Ciencias Exactas de la UNSa del 04 al 08 de octubre de 2021.

Salta, Argentina


Lic. Patricia Pesado
Coordinadora
Red UNCI


Ing. Daniel Hoyos
Decano
Facultad de Ciencias Exactas
UNSa

Formado digitalmente por: Gil, Susana Liliana
Lic. Decano
Facultad de Ciencias Exactas
Universidad Nacional de Salta



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza

ISSN: 2525-1333
Vol.: XX - Nro. XX (MES-AÑO)



Comunicación en congreso

TELEMETRÍA DE ALTAS PRESTACIONES SOBRE BASE DE DATOS DE SERIE DE TIEMPOS

HIGH PERFORMANCE TELEMETRY OVER TIME SERIES DATABASES

Pablo SOLIGO⁽¹⁾, Jorge Salvador IERACHE⁽²⁾ y Germán MERKEL⁽³⁾

⁽¹⁾Universidad Nacional de La Matanza

Correo electrónico: psoligo@unlam.edu.ar
DESEA PUBLICAR SU E-MAIL (tachar lo que no corresponda): (SI-NO)

⁽²⁾ Universidad Nacional de La Matanza

Correo electrónico: jiearche@unlam.edu.ar
DESEA PUBLICAR SU E-MAIL (tachar lo que no corresponda): (SI-NO)

⁽³⁾Universidad Nacional de La Matanza

Correo electrónico: gmerkel@unlam.edu.ar
DESEA PUBLICAR SU E-MAIL (tachar lo que no corresponda): (SI-NO)



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza



ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)

Resumen:

Los mapeadores objeto relacional (ORM) presentan grandes ventajas en términos de productividad en el desarrollo de software. Los motores de base de datos relacionales (RDBMS) representan la herramienta dominante en términos de almacenamiento y recuperación. El uso de ORMs y RDBMS sin embargo presentan límites cuando se trabaja con grandes volúmenes de datos. En este trabajo analizamos la factibilidad de implementación de una base de datos de serie de tiempos sobre un prototipo de segmento terreno costo efectivo [CITATION UGS \l 3082]. El objetivo es determinar si estas tecnologías híbridas, montadas sobre un motor relacional, permiten mantener los tiempos de almacenamiento y recuperación constantes para grandes volúmenes de datos, ofrecer capacidades ACID (Atomicidad, Consistencia, Aislamiento y Durabilidad, del inglés Consistency, Isolation and Durability) y, al mismo tiempo, compatibilidad con frameworks de desarrollo rápido y ORMS.

Abstract:

The object relational mappers (ORM) have great advantages in terms of productivity in software development. Relational database engines (RDBMS) represents the dominant tool in terms of storage and data access. The use of ORMs and RDBMS present however limits when working with large data volumes. In this paper we analyze the

feasibility of implement a time series database over a prototype of cost-effective ground segment [1]. The goal is to determine if these hybrid technologies, mounted on a relational engine, allow maintaining constant storage and recovery times for large volumes of data, offering ACID (Atomicity, Consistency, Isolation and Durability) capabilities and, at the same time, compatibility with rapid development frameworks and ORMS.

Palabras Clave: Satélites, Segmento Terreno, Diseño de Software

Key Words: Satellites, Ground segment, Software Design.

I. CONTEXTO

La UNLaM (Universidad Nacional de La Matanza) posee un prototipo de segmento terreno denominado UGS (UNLaM Ground Segment) ([CITATION UGS \l 3082], [CITATION soligo2018software \l 11274], [CITATION soligowiic2019 \l 11274] y [CITATION soligo2019software \l 11274]). En el corto plazo, bajo el proyecto PROINCE C230 el grupo de investigación GIDSA [CITATION gidsa2020 \l 11274] se ha establecido como objetivo la publicación del prototipo [CITATION UGS \l 3082] utilizando telemetría de la red satnogs [CITATION white2015satnogs \l 11274] y de otras fuentes alternativas obligando a revisar la eficiencia en el acceso a los datos. El UGS ha sido desarrollado como una plataforma de investigación, experimentación y docencia, y tiene como premisa lograr que las soluciones



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza



ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)

para el área aeroespacial sean costo-efectivas desestimando alternativas que impliquen desarrollos ad-hoc, herramientas de escasa penetración en la industria de software propósito general o de costosa implementación. Desde su primera versión, el prototipo ha tenido como objetivo ser multimisión, pudiendo trabajar con satélites de organizaciones y fabricantes independientes, lo que ha condicionado decisiones de diseño. Utilizando un RDBMS, accede a la base de datos mediante un ORM produciendo un diseño normalizado, condición que limita la eficiencia en el acceso a los datos [CITATION colley2018impact \l 11274] pero contribuye a mantener la solución costo-efectiva [CITATION russell2008bridging \l 11274].

II. INTRODUCCIÓN

Un sistema moderno de operación multimisión debe proveer de herramientas de explotación de datos que incluyan, entre otras, las siguientes características [CITATION morel2010high \l 11274] y [CITATION onlineaccessbigdataera2019 \l 11274]:

- Habilidad de mantener de forma eficiente, segura y transparente los datos de la misión completa.
- Integración para cambiar entre datos históricos y de tiempo real.
- Una performance que permita obtener datos de años en unos pocos segundos.
- Manejar grandes volúmenes de datos y niveles de variabilidad y granularidad.
- Garantía de fiabilidad.

- Punto de acceso único disponible desde interfaces de programación de aplicaciones (API del inglés Application Programming Interface)

Estas características se vuelven especialmente importantes ante los avances disponibles en el área de aprendizaje automático, un área que exige alta disponibilidad y grandes volúmenes de datos históricos [CITATION ibrahim2018machine \l 11274] para su explotación.

Los desafíos que presenta la administración de grandes volúmenes de datos no son nuevos, aunque en las últimas décadas el surgimiento de las redes sociales, aplicación de uso masivo en internet y la popularidad adquirida por dispositivos interconectados IoT(Internet de las cosas, del inglés Internet of Things) han propiciado la aparición de herramientas especialmente adaptadas. Las soluciones genéricamente denominadas NoSql (No solo Sql, del inglés Not Only Sql) intentan ofrecer una solución escalable al problema que presenta las, cada vez mayores, necesidades de almacenamiento y posterior recuperación.

Por otro lado los sistemas RDBMS clásicos han dominado casi desde de su aparición y lo siguen haciendo concentrando más del 80% de las soluciones de almacenamiento y recuperación [CITATION serra2015relational \l 11274], manteniendo su popularidad como muestra la figura 1 Popularidad de modelos de base de datos. El éxito de este tipo de almacenamiento tiene entre sus razones, que los datos empresariales ajustan al modelo relacional [CITATION serra2015relational \l 11274]. Un segmento terreno aeroespacial concentra ambas problemáticas, grandes



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza



ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)

volúmenes de datos con característica que se adaptan perfectamente al modelo relacional.

Las características de seguridad ACID [CITATION mohamed2014relational \l 11274] ofrecidas por un RDBMS, ausentes o parcialmente disponibles en un NoSql, también juegan un papel en la determinación del modelo de base de datos a implementar.

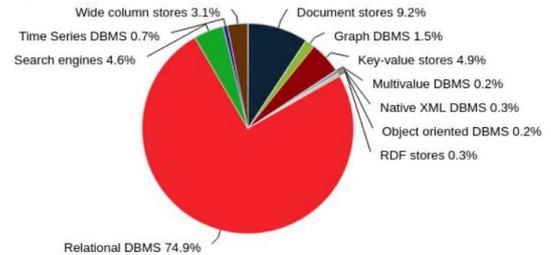


Fig 1 Popularidad de modelos de base de datos
Como muestra la Fig 2 Tabla de variables de ingeniería y relacionadas, el UGS posee una única tabla donde almacenar las variables de ingeniería, este diseño de tipo Narrow-table [CITATION mcbride2020survey \l 11274] y [CITATION TimeScaleDBDocumentation \l 11274], es simple y a la vez lo suficientemente versátil para adaptarse a escenarios con diferente cantidad y tipos de sensores aunque afecta negativamente el rendimiento ya que cada registro almacena un único valor de telemetría. Bajo este diseño y utilizando RDBMS algunas entidades pueden sufrir problemas de escalabilidad si la cantidad de datos crece demasiado, problemas relacionados con las limitaciones en bases de datos estrictamente normalizadas.

Las soluciones a los límites impuestos por la tecnología pueden ir desde desnormalizaciones que incluyan vistas materializadas con datos conciliados, bases de datos NoSQL o soluciones híbridas que implementen capas lógicas adicionales a probados RDBMS [CITATION distbigastronomicalcatalogues2019 \l 11274].

Las desnormalizaciones sobre RDBMS, han demostrado ser exitosas en varios ámbitos de la industria y los segmentos terrenos no son la excepción. Grandes compañías lo han usado como estrategia [CITATION morel2010high \l 11274] teniendo como una de sus grandes ventajas el uso de una única, probada y madura



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



herramienta de persistencia y recuperación, evitando el uso de herramientas especiales para datos según su naturaleza. Como contrapartida, un diseño desnormalizado no puede por sí mismo garantizar la integridad, y conlleva un costo asociado a la implementación de lógica de control y actualización adicional.

Las soluciones NoSQL, o el uso de motores de persistencia que no posean características ACID debería circunscribirse un grupo de entidades particulares, aquellas que presentan características especiales ya sea de acceso, actualización o volumen de datos. En cualquier caso, esta estrategia obliga a utilizar una herramienta adicional, generando costos adicionales de mantenimiento y desarrollo.

El surgimiento en los últimos años de alternativas híbridas, adaptadas a problemáticas de administración de datos específicas, que ofrezcan mejores rendimientos sin perder la seguridad y estandarización presentes en las RDBMS ha ganado atención incluso en ámbitos de máxima exigencia [CITATION stefancova2018evaluation \ 11274].

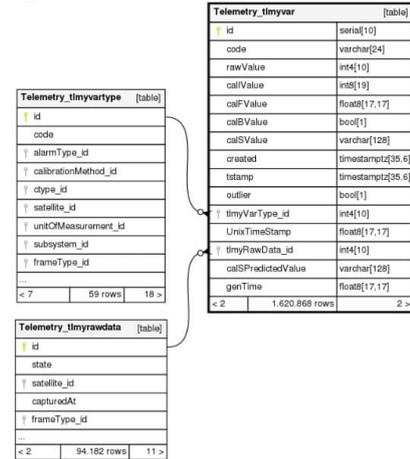


Fig 2 Tabla de variables de ingeniería y relacionadas

En este trabajo exponemos los resultados obtenidos de las simulaciones realizadas para evaluar las capacidades de estos sistemas híbridos sobre las necesidades de un segmento terreno de bajo costo. El objetivo es evaluar la viabilidad de utilizar esta tecnología que permitiría una migración transparente, costos de desarrollo y mantenimiento acotados y poder ofrecer rendimientos constantes de manera independiente a la cantidad de misiones y la cantidad de datos de cada una.

III. MÉTODOS

Con el objetivo de verificar las posibles diferencias de rendimiento y la escalabilidad de un sistema híbrido se ha utilizado la extensión de PostgreSQL, TimescaleDB [CITATION TimeScaleDB \ 11274] aplicada a la base de datos actual del UGS. No se han aplicado modificaciones al diseño buscando que continúe siendo accedida tanto como sea posible mediante un ORM.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Para las pruebas y simulación se ha tomado como referencia 3 satélites simulados con 7500 variables de telemetría con una frecuencia de reporte de 8 segundos. Utilizando la API de timescaleDB *create\hypertable* se generó una hipertabla *Telemetry\htlmyvar* replica de la tabla relacional *Telemetry\tlmyvar* y se han configurado periodos de tiempo en 2 horas buscando hacer coincidir el tamaño de los fragmentos *chunks* con el tamaño recomendado por el equipo de desarrollo de *timescaledb* [CITATION TimeScaleDBDocumentation \l 11274] y un periodo de tiempo discreto. A la tabla regular se le ha agregado un índice ordenado, no cluster, por fecha hora. Con esto último se intenta dar a la tabla regular herramientas similares que las que la hipertabla implementa.

La Tabla 1 - Hardware utilizado en las simulaciones, muestra una lista del equipamiento de hardware utilizado.

Tabla 1 - Hardware utilizado en las simulaciones

Material	Descripción
SO	UBUNTU
DBMS	Timescale sobre Docker
Equipo	
Procesador	AMD A10-7860K Radeon R7, 12 Compute Cores 4C+8G
Memoria	6916MiB
Disco	Western Digital 6TB SATA 3.5 Hard Drive.

La siguiente lista describe las simulaciones realizadas.

- Inserciones: Inserciones de telemetría semiordenada
- Consultas de tiempo real: 250 consultas desde entre 10 y 50 minutos antes de la fecha/hora del

último registro hasta la fecha hora del último registro. Las consultas son realizadas por 5 procesos.

- Consultas históricas: 250 consultas desde entre 10 y 50 minutos en todo el rango de datos. Las consultas son realizadas por 5 procesos.
- Consultas agregadas: 250 consultas desde entre 10 y 50 minutos en todo el rango de datos agrupando cada 12 minutos.

Para las inserciones se genera de manera simulada telemetría para 4 paquetes para cada satélite, posteriormente se mezclan para evitar la inserción en orden y lograr un escenario más realista. Finalmente se insertan mediante el ORM en bloques de 1000.

IV. RESULTADOS

Las figuras Fig 3 Inserciones entre los 450 y 550 millones de registros y Fig 4 Inserciones entre los 1010 y 1060 millones de registros muestran un comportamiento similar e independiente a la cantidad de registros. Se observan mejores rendimientos para la hipertabla. La tabla regular muestra valores altos periódicamente para una inserción mientras que la hipertabla muestra un comportamiento totalmente estable.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de Ingeniería e Investigaciones Tecnológicas de la Universidad Nacional de La Matanza



ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)

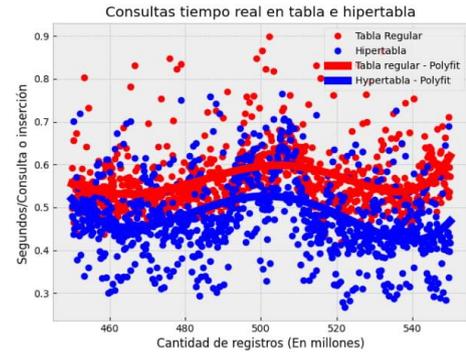
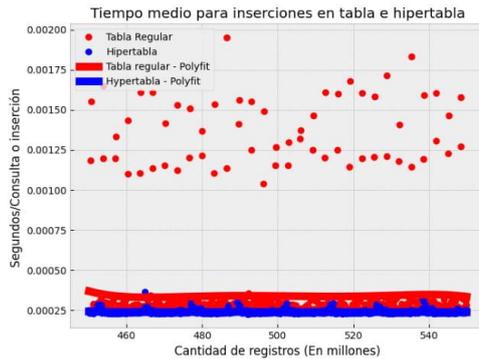


Fig 3 Inserciones entre los 450 y 550 millones de registros

Fig 5 Consultas de tiempo real entre 450 y 550 millones de registros

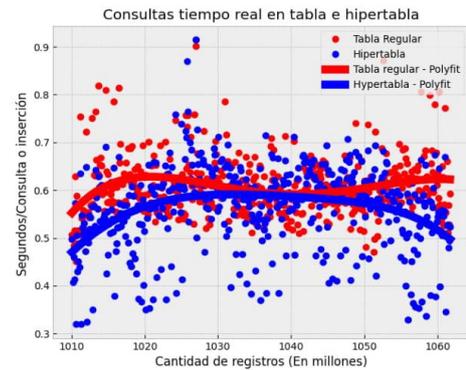
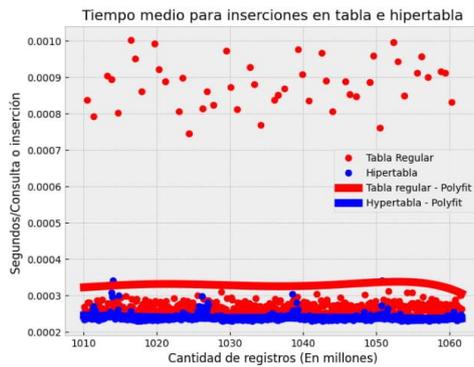


Fig 4 Inserciones entre los 1010 y 1060 millones de registros

Fig 6 Consultas en tiempo real entre 1010 y 1060 millones de registros

Las Fig 5 Consultas de tiempo real entre 450 y 550 millones de registros y Fig 6 Consultas en tiempo real entre 1010 y 1060 millones de registros muestran un mejor rendimiento para la hipertabla, las consultas demoran de media un tiempo inferior que trabajando con tablas regulares.

Las figuras Fig 7 Consultas históricas entre 450 y 550 millones de registros y Fig 8 Consultas históricas entre 1010 y 1060 millones de registros muestran, en contraste con las figuras previas, un mejor rendimiento para la tabla regular, llegando incluso a reducir a la mitad el tiempo de consulta.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza



ISSN: 2525-1333
Vol.: XX - Nro. XX (MES-AÑO)

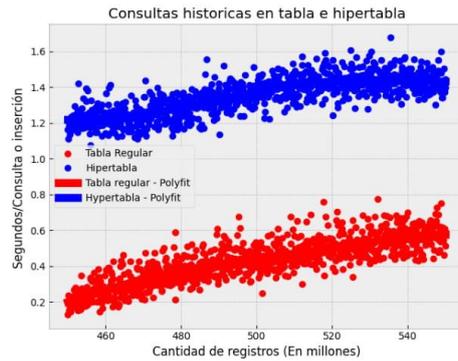


Fig 7 Consultas históricas entre 450 y 550 millones de registros

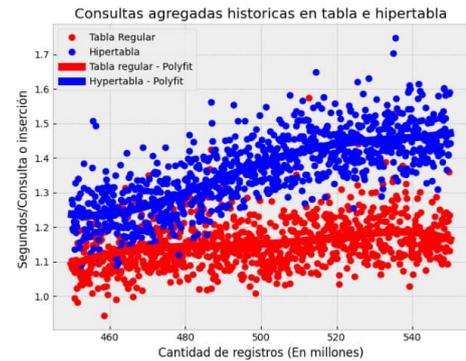


Fig 9 Consultas agregadas entre 450 y 550 millones de registros

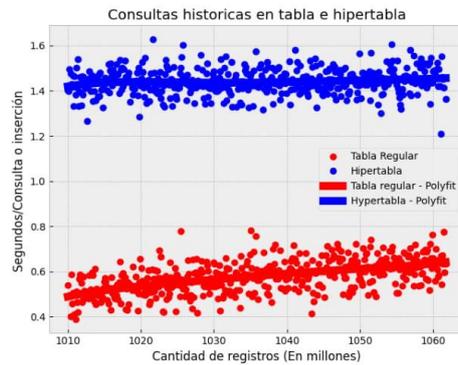


Fig 8 Consultas históricas entre 1010 y 1060 millones de registros

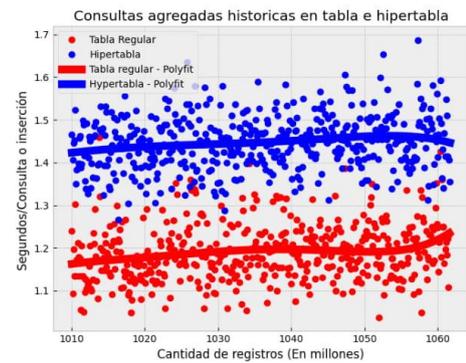


Fig 10 Consultas agregadas entre 1010 y 1060 millones de registros

Al igual que con las consultas históricas las figuras Fig 9 Consultas agregadas entre 450 y 550 millones de registros y Fig 10 Consultas agregadas entre 1010 y 1060 millones de registros muestran un mejor rendimiento sobre tablas regulares que sobre hipertablas.

VI. CONCLUSIONES

Los resultados muestran que las inserciones son más eficientes en todos los casos sobre la hipertabla. Para las consultas el escenario es mixto con mejores rendimientos



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza

ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)



en uno u otro caso según la naturaleza de la consulta. Hasta la cantidad de registros simulados (1100 millones) ambas estrategias se mostraron estables, no hay una degradación marcada en ninguna de ellas ni tampoco rendimientos que ofrezcan una clara ventaja en esta instancia. Si bien las inserciones pueden ser un cuello de botella, por su naturaleza, el sistema espera recibir muchas consultas principalmente de carácter histórico. Este escenario no deja claro que una opción ofrezca mejores resultados generales por sobre la otra. Con los datos obtenidos a la fecha se continuará trabajando sobre una versión postgresql estándar. No se descarta a futuro la migración a timescale pero la misma se considera injustificada según las simulaciones realizadas.

VII. REFERENCIAS Y BIBLIOGRAFÍA

A. Referencias bibliográficas:

- [1] <https://ugs.unlam.edu.ar>, UNLaM Ground Segment, 2020.
- [2] P. Soligo y J. S. Ierache, «Software de segmento terreno de próxima generación,» de *XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018)*, 2018.
- [3] P. Soligo y J. S. Ierache, «Segmento Terreno Para Misiones Espaciales de Próxima Generación,» *WICC 2019*.
- [4] P. Soligo y J. S. Ierache, «Arquitectura de segmento terreno satelital adaptada para el control de límites de telemetría dinámicos,» de *XXV Congreso Argentino de Ciencias de la Computación (La Plata, 2019)*, 2019.
- [5] <https://gidsa.unlam.edu.ar>, *Grupo de Investigación y Desarrollo de Software Aeroespacial de la*



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza



ISSN: 2525-1333

Vol.: XX - Nro. XX (MES-AÑO)

Universidad Nacional de La Matanza, 2020.

- [6] D. J. White, I. Giannelos, A. Zissimatos, E. Kosmas, D. Papadeas, P. Papadeas, M. Papamathaiou, N. Roussos, V. Tsiligiannis y I. Charitopoulos, «SatNOGS: satellite networked open ground station,» 2015.
- [7] D. Colley, C. Stanier y M. Asaduzzaman, «The Impact of Object-Relational Mapping Frameworks on Relational Query Performance,» de *2018 International Conference on Computing, Electronics & Communications Engineering (ICCECE)*, 2018.
- [8] C. Russell, «Bridging the object-relational divide,» *Queue*, vol. 6, pp. 18-28, 2008.
- [9] T. Morel, G. Garcia, M. Palsson y J. C. Gil, «High Performance Telemetry Archiving and Trending for Satellite Control Centers,» de *SpaceOps 2010 Conference Delivering on the Dream Hosted by NASA Marshall Space Flight Center and Organized by AIAA*, 2010.
- [10] G. Pace, M. Schick, A. Colapicchioni, A. Cuomo y U. Voges, «EO ON-LINE DATA ACCESS IN THE BIG DATA ERA,» de *Proceedings of the 2019 conference on Big Data from Space*, 2019.
- [11] S. K. Ibrahim, A. Ahmed, M. A. E. Zeidan y I. E. Ziedan, «Machine learning methods for spacecraft telemetry mining,» *IEEE Transactions on Aerospace and Electronic Systems*, vol. 55, pp. 1816-1827, 2018.
- [12] J. Serra, «Relational databases vs Non-relational databases,» *Viitattu*, vol. 16, p. 2017, 2015.
- [13] M. A. Mohamed, O. G. Altrafi y M. O. Ismail, «Relational vs. nosql databases: A survey,» *International Journal of Computer and Information Technology*, vol. 3, pp. 598-601, 2014.
- [14] B. McBride y D. Reynolds, «Survey of time series database technology,» 2020.
- [15] <https://docs.timescale.com/>, *TimescaleDB Documentation*, 2017.
- [16] S. N. Pilar de Teodoro y J. Salgado, «Distributing big astronomical catalogues with Greenplum,» de *Proceedings of the 2019 conference on Big Data from Space*, 2019.
- [17] E. Stefancova, «Evaluation of the TimescaleDB PostgreSQL Time Series extension,» 2018.
- [18] <https://www.timescale.com/>, *TimescaleDB*, 2017.
- [19] A. Jacobs, «The pathologies of big data,» *Queue*, vol. 7, pp. 10-19, 2009.
- [20] A. B. M. Moniruzzaman y S. A. Hossain, «Nosql database: New era of databases for big data analytics-classification, characteristics and comparison,» *arXiv preprint arXiv:1307.0191*, 2013.
- [21] G. Graefe y J. Alger, *Electronic database operations for perspective transformations on relational tables using pivot and unpivot columns*, Google Patents, 2001.
- [22] blog.timescale.com, *Time-series data: Why (and how) to use a relational database instead of NoSQL*, 2017.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



Revista Digital del Departamento de
Ingeniería e Investigaciones
Tecnológicas de la Universidad
Nacional de La Matanza

ISSN: 2525-1333
Vol.: XX - Nro. XX (MES-AÑO)



Recibido: A completar por el Editor. Formato: AAAA-MM-DD
Aprobado: A completar por el Editor. Formato: AAAA-MM-DD
Hipervínculo Permanente: A completar por el Editor
Datos de edición: Vol. [A completar por el Editor]-Nro. [A completar por el Editor]-Art. [A completar por el Editor]
Fecha de edición: Formato: AAAA-MM-DD





Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Investigación, desarrollo y publicación de un prototipo de segmento terreno satelital

Pablo Soligo , German Merkel, Jorge S. Ierache 

Grupo de Investigación y Desarrollo de Software Aeroespacial (GIDSA)
Departamento Ingeniería e Investigaciones Tecnológicas (DIIT)
Universidad Nacional de la Matanza (UNLaM)

psoligo@unlam.edu.ar; gmerkel@alumno.unlam.edu.ar; jierache@unlam.edu.ar

RESUMEN

Las líneas de investigación y desarrollo presentadas tienen por objetivo demostrar la factibilidad de desarrollar sistemas de segmento terreno satelital costo-efectivos utilizando exclusivamente componentes dentro de los denominados, de estantería, en cualquiera de sus variantes OTS (Del Inglés Off-the-Shelf), COTS (Del Inglés Commercial-Off-The-Shelf) y OSS (Del Inglés Open Source Software) prescindiendo de soluciones propias o de herramientas de escasa penetración en la industria de software de propósito general.

Palabras clave: *Segmento Terreno, Software, Costo-Efectivo.*

CONTEXTO

Las experiencias realizadas en la Maestría en Desarrollos Informáticos de Aplicación Espacial (MDIAE) (Comisión Nacional de Actividades Espaciales-Universidad Nacional de La Matanza), tanto de manera directa, operando unidades de software de segmento terreno de varias agencias, como mediante investigación general de las soluciones implementadas en el área, propiciaron la creación del grupo de investigación y desarrollo de software aeroespacial de la Universidad Nacional de La Matanza (GIDSA), en el marco de proyecto “Proince C-211: Sistemas de segmento terreno satelital de próxima generación” y actualmente se radica en el proyecto “Proince C-230, Aprendizaje automático para el control del estado de salud en sistemas Aeroespaciales”. El GIDSA [1] está

dedicado a investigar e implementar prototipos de software alternativos de bajo costo basados en las soluciones ampliamente aceptadas, de probada madurez y con penetración en la industria de software de propósito general.

1. INTRODUCCIÓN

El alto costo asociado a las misiones espaciales y la baja propensión a tomar riesgos determina el enfoque en las soluciones implementadas en la industria espacial ([2], [3]). El desarrollo de software está fuertemente orientado al cumplimiento de los requerimientos dificultando una estrategia más amplia y de visión de largo plazo. El grupo de investigación GIDSA desarrolla un prototipo de segmento terreno alternativo genérico, costo efectivo y basado completamente en COTS, OSS o OTS, minimizando no solo el desarrollo sino también el costo de mantenimiento. La solución que el grupo GIDSA propone una alternativa que prescinde de software propietario o de propósito específico, haciendo a esta misma portable y poderosa.

2. LÍNEAS DE INVESTIGACIÓN Y DESARROLLO

Las líneas de investigación en desarrollo proponen la exploración de alternativas basadas exclusivamente en técnicas y herramientas de alta penetración en la industria del software, en particular, aquellas que puedan ser aplicadas en la industria espacial. El objetivo principal es demostrar que las herramientas usadas en la industria de software de propósito general pueden ser aplicadas en la industria espacial, no solamente



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

en el segmento terreno sino también en el segmento de vuelo. La aplicación de estas técnicas puede ayudar a desarrollar soluciones de bajo costo con un alto nivel de mantenibilidad.

Capa de visualización y operación

La capa de visualización del Unlam Ground Segment (UGS) está basada en el NASA OpenMCT (Open Mission Control Technologies). El NASA OpenMCT es el framework de visualización de control de misión de código abierto de la NASA.

Se evaluaron múltiples alternativas, partiendo de la necesidad de tener acceso desde múltiples plataformas. Las soluciones nativas fueron descartadas al requerir un desarrollo específico por plataforma y no mostrarse como una alternativa costo efectiva. Se han explorado soluciones basadas en paneles de comandos, como Grafana (<https://grafana.com/>). Estas alternativas resuelven de manera satisfactoria la visualización, pero son pretendidamente genéricas y carecen de necesidades propias de la industria aeroespacial. La actualización en tiempo real, paneles de comandos, si bien son técnicamente posibles, su implementación no resulta transparente, siendo estas necesidades en un software de control de misión.

Por otro lado, las soluciones SPA se presentan como la alternativa tecnológica que permite un único desarrollo compatible con prácticamente cualquier plataforma, ofreciendo un nivel de interactividad comparable, en determinados escenarios, con una aplicación nativa. Se exploró un desarrollo SPA propio invirtiéndose más de 120 horas alcanzando pobres resultados [4].

Finalmente, la implementación de la capa de visualización sobre OpenMCT fue considerada la mejor opción, ofreciendo una SPA con funcionalidad específica del área espacial y un costo de adaptación inferior al costo de desarrollo de una SPA propia.

Las interfaces del Unlam Ground Segment (UGS) están completamente basadas en la comunicación HTTP/HTTPS, siendo este un estándar en la industria del software. Toda la telemetría es incorporada y puede ser extraída del sistema mediante un servicio REST lo que permitió una integración transparente con el OpenMCT

El prototipo actual (Disponible en <https://ugs.unlam.edu.ar>), [5] permite visualizar telemetría de distintas formas, tales como gráficos y tablas, crear múltiples tableros y paneles, guardar y compartir. La visualización se conecta vía HTTP y JSON modularmente, permitiendo añadir nuevos satélites o nuevas variables de telemetría sin necesidad de modificar la aplicación web en sí. OpenMCT provee al desarrollador un framework con tableros predeterminados, almacenamiento local en el navegador y un módulo de complementos.

Recuperación y persistencia

La definición de datos como alarmas, variables de telemetría, comandos, tipos de datos y formatos son almacenados sobre un RDBMS. El UGS utiliza desde su primera versión un RDBMS con el objetivo de estandarizar el almacenamiento, recuperación, seguridad e integridad de los datos en cualquier nivel de procesamiento. Los datos son accedidos por medio de un ORM, esta capa intermedia provee productividad en el desarrollo e independencia del proveedor del motor de base de datos, haciendo sencillo, dentro de ciertos límites, el cambio de RDBMS.

Por otro lado, el uso de ORMs y RDBMS, con modelos estrictamente normalizados puede establecer un límite en el rendimiento cuando se trabaja con grandes volúmenes de datos. Las tecnologías híbridas, montadas sobre motores relacionales, han sido testeadas en busca de mejores rendimientos tanto en el almacenamiento como en la recuperación para grandes volúmenes de datos.

Detección de fallas

Existen tres métodos comunes de análisis de telemetría y control de salud: control de límites, sistemas expertos y sistemas basados en modelos. El primer método es el más simple y el más común del grupo, consiste en establecer un valor aceptable máximo y un valor aceptable mínimo para un sensor con la ayuda de un experto, y verificar que los valores de telemetría de dicho sensor se mantengan entre los valores definidos. Este método es completamente insensible al contexto y puede ser tedioso establecer valores límites a cada sensor. El segundo método obtiene mejores resultados, aunque no puede encontrar tipos de fallas no consideradas o indefinidas, y requiere una laboriosa



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

configuración. La idea detrás del tercer método es detectar anomalías y sus razones comparando simulaciones computacionales con el verdadero comportamiento del sistema.

Los tres métodos requieren que un experto continuamente complete ciertas tareas, tales como actualizar límites, crear reglas o parametrizar situaciones. Por otro lado, el aprendizaje automático provee un largo rango de posibilidades para la predicción de comportamientos, y, por tanto, detección de fallas. En lugar de un experto infringiendo reglas y desarrollando modelos, el aprendizaje automático puede ofrecer una manera más eficiente de capturar conocimiento y aplicarlo. Desde el comienzo, el UGS verifica la salud de un satélite utilizando control de límites tanto como para variables directas como para variables derivadas. Trabajos anteriores [7] muestran la creación de una entidad asociada al tipo de telemetría donde un modelo de predicción es automáticamente creado, por tanto, presentando las habilidades del aprendizaje automático. Estableciendo dinámicamente los valores máximos y mínimos admisibles utilizando la predicción obtenida, se puede lograr que el sistema obtenga controles actualizados sensibles al contexto para el satélite.

El caso de estudio cubre el análisis de un sensor de tensión de baterías durante un eclipse, teniendo en cuenta el tiempo que el satélite se encuentra eclipsado y la tensión de la batería. La tensión de la batería normalmente decrece durante un eclipse dado que los paneles solares del satélite se encuentran parcial o totalmente tapados por la Tierra. Una vez que los paneles solares enfrentan el Sol, la tensión de la batería se recupera rápidamente.

Usando el método de control de límites, los valores mínimos y máximos deben tener en cuenta el eclipse, y, por tanto, la amplitud de estos debe ser lo suficientemente grande. Sin embargo, existe la posibilidad de que el satélite no se encuentre eclipsado por la Tierra y que su batería se encuentre con baja tensión. Dado que el método de control de límites no conoce cuándo el satélite está o no eclipsado, no detectará ninguna anomalía (aunque en realidad puede haber un serio problema con la carga de la batería). Un grupo de reglas proveídas

por un experto puede mejorar esta situación, aunque en este caso un experto es necesario y los valores mínimos y máximos finales son tan solo conocidos con el sistema en vuelo.

Con el aprendizaje automático, un proceso distribuido prueba los tipos de telemetría que tienen modelos de predicción expirados. Por cada tipo de telemetría, el sistema crea una nueva predicción de datos acorde a su configuración, crea un nuevo modelo de predicción y, en caso de ser lo suficientemente preciso, persiste el modelo, por el contrario, se señala una alarma.

Se representa en la Figura 1 - Arquitectura Conceptual muestra los principales módulos presentes en el prototipo.

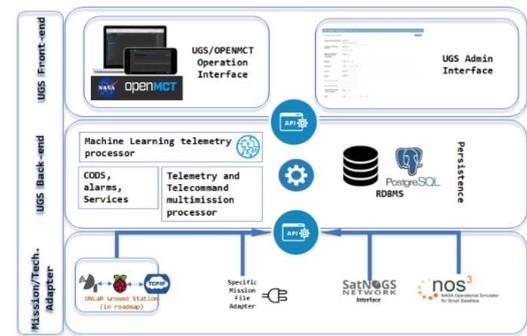


Figura 1 - Arquitectura Conceptual

La Figura 2 muestra un dashboard ejemplo con datos del satélite de satellogic butsat1 (Tita). Accesible en <https://ugs.unlam.edu.ar/#/browse/mine/demodash>



Figura 2 – UGS/NASA OPENMCT Front-end



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLAM
Versión	5
Vigencia	03/9/2019

3. RESULTADOS OBTENIDOS/ESPERADOS

El UGS ha sido publicado [8] en el año 2020, donde la telemetría de los satélites BugSat1/TITA y CSIM-FD puede ser explorada y analizada. La telemetría de dichos satélites ha sido provista mayormente por la red SatNOGS. La actual capa de visualización provee una manera interactiva de explorar la telemetría de ambos satélites de manera histórica y parcialmente en tiempo real. Los paquetes son decodificados usando el procesador de telemetría descripto. Características específicas son procesadas utilizando scripts programados en lenguajes de propósito general, cargados en tiempo de ejecución [3]. Se espera añadir a la capa de visualización del UGS una interfaz completa de telemetría en tiempo real y de envío de comandos. Los scripts de comandos serán desarrollados, como se ha dicho, un lenguaje de programación de propósito general. Aunque el OpenMCT sea un framework completo, se han requerido no menos de 100 horas de desarrollo para lograr una integración parcial. Se requiere entrenamiento en el lenguaje, herramientas y estructuras necesarias para la adaptación. El framework está bien documentado, pero carece de tutoriales útiles y ejemplos, y su desarrollo puede ser complejo. Con todo, OpenMCT fue considerado la mejor opción dado el alto costo del desarrollo de SPA: problemas que el framework soluciona parcialmente.

En términos de persistencia el uso de soluciones híbridas no han mostrado una clara ventaja que justifique su implementación [6].

Las primeras implementaciones de aprendizaje automático aplicado a la telemetría satelital han dado resultados prometedores [3] aunque no está exenta de problemas o de puntos de investigación por explotar. La búsqueda de relaciones o las variables de contexto todavía requieren del asesoramiento de un experto para su correcta identificación. Una de las futuras líneas de investigación es la implementación de una solución al problema de encontrar correlaciones y la corrección automática de estas, esperando utilizar completamente el poder computacional, no para reemplazar al experto, sino para encontrar en el un aliado en la búsqueda e identificación de patrones escondidos.

4. FORMACIÓN DE RECURSOS HUMANOS

Los prototipos desarrollados presentan una plataforma realista de experimentación. Le permite a investigadores y estudiantes probar soluciones de software, obtener límites, comparar alternativas y establecer criterios de decisión. La posibilidad de trabajar con datos de varias misiones espaciales desde pequeñas misiones universitarias hasta grandes misiones científicas permite responder a la premisa de desarrollar un sistema terreno transparente al satélite en órbita. Actualmente el grupo de investigación este compuesto por un investigador formado, un investigador en formación y un alumno investigador becario BIC (Beca de investigación científica UNLAM).

5. BIBLIOGRAFÍA

- [1] <https://gidsa.unlam.edu.ar/>, *Grupo de Investigación y Desarrollo de Software Aeroespacial de la Universidad Nacional de La Matanza*, 2020.
- [2] P. Soligo y J. S. Ierache, «Segmento Terreno Para Misiones Espaciales de Próxima Generación,» *WICC 2019*.
- [3] P. Soligo y J. S. Ierache, «Software de segmento terreno de próxima generación,» de *XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018)*, 2018.
- [4] P. Soligo y J. S. Ierache, «Experiences and lessons learned developing a next-generation ground segment prototype,» *2nd IAA Latin American Symposium on Small Satellites*.
- [5] *openmct nasa open mission control technologies*.
- [6] P. Soligo, G. Merkel y J. S. Ierache, «Telemetría de altas prestaciones sobre base de datos de serie de tiempos,» *REDDI 2021*, 2021.
- [7] P. Soligo y J. S. Ierache, «Arquitectura de segmento terreno satelital adaptada para el control de límites de telemetría dinámicos,» de *XXV Congreso Argentino de Ciencias de la Computación (La Plata, 2019)*, 2019.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

[8] <https://ugs.unlam.edu.ar/>, *Unlam Ground Segment*, 2020.

[9] J. J. Ramos Pérez, «A design for an advanced architecture of satellite ground segments,» 2014.

ACRÓNIMOS

COTS	Commercial-Off-The-Shelf..... 1
GIDSA	Grupo de Investigación y Desarrollo de Software Aeroespacial. 1
HTTP	Protocolo de transferencia de hipertexto, del Inglés Hypertext Transfer Protocol 2
HTTPS	Protocolo de transferencia de hipertexto seguro, del inglés Hypertext Transfer Protocol Secure..... 2
NASA	Administración Nacional de Aeronáutica y el Espacio, del Inglés, National Aeronautics and Space Administration 2, 4

OpenMCT	Software de control de misión abierto, del Inglés Open Source Mission Control Software..... 2
ORM	Mapeo objeto-relacional, del inglés Object-Relational Mapping2
OSS	Open Source Software 1
OTS	Off-the-Shelf 1
RDBMS	Sistema de administración de base de datos relacional, del Inglés Relational Database Management System 2
REST	Transferencia de Estado Representacional, del inglés representational state transfer 2
SatNOGS	Satellite Networked Open Ground Station 4
SPA	Aplicación de página única, del Inglés Single Page Application . 2, 4
UGS	UNLaM Ground Segment..... 2, 3, 4



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Detección de Anomalías en Segmento Terreno Satelital Aplicando Modelo de Mezcla Gaussiana y Rolling Means al Subsistema de Potencia.

Pablo Soligo, Germán Merkel, and Jorge Ierache

Universidad Nacional de La Matanza,
Florencio Varela 1903 (B1754JEC) San Justo, Buenos Aires, Argentina
{psoligo, jierache}@unlam.edu.ar
{gmerkel}@alumno.unlam.edu.ar
<http://unlam.edu.ar>

Abstract. En este trabajo exploramos la posibilidad de encontrar anomalías automáticamente en telemetría satelital real. Comparamos dos técnicas de aprendizaje automático diferentes como alternativa al control de límites clásico. Intentamos evitar, en la medida de lo posible, la intervención de un experto, detectando anomalías que no se pueden encontrar con los métodos clásicos o que se desconocen de antemano. La mezcla gaussiana y Rolling Means se aplican en la telemetría del subsistema de potencia de un satélite órbita baja. Algunos valores de telemetría se modificaron artificialmente para generar un apagado en un panel solar para intentar lograr una detección temprana por contexto o por comparación. Finalmente, se presentan los resultados y la conclusión.

Keywords: Satellites, Ground Segment, Platform, Telemetry, Machine Learning, Data Mining, Anomaly Detection

1 Introducción

El Grupo de Investigación y Desarrollo de Software Aeroespacial (GIDSA) ?? tiene como objetivo proponer y probar prototipos de soluciones de software para el área aeroespacial de nueva generación. El trabajo desarrollado incluye prototipos que utilizan interpretes de propósito general para decodificar telemetría y scripts de comandos, adopción de estándares bien probados en la industria del software, almacenamiento masivo de telemetría y detección de fallas [1], [2] y [3]. El prototipo funcional del segmento terreno se encuentra público en internet y funciona con datos de satélite reales, principalmente obtenidos de la red SatNOGS [4] y [5].

La detección temprana de anomalías en sistemas complejos como satélites artificiales son de vital importancia teniendo en cuenta el costo de las misiones y la dificultad de reparar daños. El control de límites superior e inferior para muchas variables de telemetría suelen ser la técnica más común para detectar comportamientos anómalos[6]. Como se indicó en un artículo anterior [7], la



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

salud del satélite se controla con la ayuda constante de un experto, utilizando poca potencia computacional. Mientras tanto, en la industria del software, el aprendizaje automático se utiliza actualmente para diferentes tipos de detección de anomalías, como fraudes con tarjetas de crédito y detección de intrusiones entre otros [8] y [9]. El objetivo de utilizar el aprendizaje automático es lograr una detección temprana de fallas evitando, en la medida de lo posible, la evaluación constante por parte de expertos así como detectar tipos de anomalías desconocidas previamente. El aprendizaje automático ofrece una interesante variedad de posibilidades de predicción y detección de anomalías. Hay dos tipos de algoritmos de aprendizaje automático: aprendizaje automático supervisado y no supervisado. El primero depende de los datos de entrada etiquetados, es decir, el conjunto de datos de entrada debe haber definido si un dato se considera una anomalía o no. El aprendizaje no supervisado no depende de los datos de entrada etiquetados, sino que aprende la representación interna del conjunto de datos y genera patrones [10]. Una anomalía es cualquier dato que se desvía de lo esperado o normal. En la literatura estadística, también se les conoce como valores atípicos o outliers. Cada dato que es procesado por el prototipo será clasificado usando etiquetas binarias: un dato es una anomalía o no [8]. Para detectar anomalías, los algoritmos de aprendizaje automático de detección de fallas crean un modelo del patrón nominal en el conjunto de datos, luego calculan una puntuación para cada valor como medida de cuán atípico es. Dependiendo del algoritmo, esta puntuación atípica toma en cuenta la correlación con diferentes características o no [8]. En este trabajo y en el caso de series de datos de tiempo, buscamos una secuencia de valores atípicos que determina una anomalía en lugar de un dato particular, buscamos un comportamiento anormal del sistema en lugar de un valor incorrecto.

El UNLaM Ground Segment (UGS) posee el control de límites clásico desde su primera versión. En versiones posteriores se implementaron módulos prototipo que modifican los límites dinámicamente[7]. En este trabajo, en lugar de trabajar con límites, buscamos obtener una medida de éxito en la detección de un comportamiento anómalo del subsistema de potencia, comportamiento que no puede ser detectado por el control de límites clásico. El trabajo actual se exploran dos métodos de aprendizaje automático diferentes mezcla gaussiana y rolling means. Estos dos algoritmos son investigados y comparados entre sí para estudiar la viabilidad de aplicarlos en la detección de patrones y comportamientos en un prototipo de control de salud en tiempo real.

2 Materiales y métodos

Para estos experimentos usamos nuestro propio conjunto de datos con datos de telemetría real [11]. La fuente de la telemetría es un satélite científico de órbita baja. La telemetría comienza en 2015-05-27 08:51:06 +00:00 y termina en 2015-06-05 23:34:06 +00:00. Para estos experimentos usamos solo los dos primeros días, desde 2015-05-27 08:51:22 +00:00 hasta 2015-05-29 08:50:59 +00:00. El conjunto de datos de entrenamiento finalmente tiene 17277 tuplas, el conjunto



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

de datos de prueba tiene 4320 tuplas. La tabla 1 muestra el significado de cada campo según descripción disponible en la documentación del fabricante. Todos los valores, excepto *vBatAverage* y *BatteryDischarging* están en bruto(raw), sin embargo, los datos siempre se normalizan antes de ser procesados. Desafortunadamente, el conjunto de datos no cuenta con fallas documentadas por un experto. Para crear una anomalía artificial, cortamos parcialmente la generación energía del panel solar 24 poniendo en 0 la corriente (128 en bruto) en el conjunto de datos de prueba. El corte es progresivo y cubre 1079 tuplas. Esto es similar a dejar el panel eclipsado (según posición orbital), independientemente del contexto real. Tenga en cuenta que el límite clásico el control no puede manejar este comportamiento, debido al hecho de que las corrientes cercanas a 0 son perfectamente válidas en períodos de eclipse.

Feature	Meaning
vBatAverage	Average of Battery voltage used by supervisions
BatteryDischarging	Flag True/False if battery is discharging
ISenseRS1	ISenseRS1 current (battery current)
ISenseRS2	ISenseRS2 current (battery current)
V_MODULE_N_SA	Current in solar panel #N con $0 < N < 25$

Table 1: DataSet Features

Se utilizan dos algoritmos de aprendizaje automático diferentes para detectar anomalías en Telemetría satelital: Mezcla gaussiana y Rolling Means. El primero se aplica la telemetría del subsistema de potencia en su conjunto, utilizando la correlación entre variables, mientras que la última se aplica a cada variable de telemetría de manera aislada. Ambos modelos siguen enfoques estadísticos clásicos: ambos utilizan medidas estadísticas como media, desviación estándar y probabilidad.

2.1 Gaussian Mixture Model

Todas las telemetrías del subsistema de potencia están altamente correlacionadas como se muestra en la figura 1, por razones de tamaño, mostramos la correlación de solo 4 de las 24 características actuales del panel.

Usando la biblioteca sklearn [12], creamos una Modelo de Mezcla Gaussiana, del Inglés Gaussian Mixture Model (GMM). GMM puede utilizarse para agrupar datos sin etiquetar, GMM puede ayudar a detectar un comportamiento lejano o poco probable que el comportamiento nominal. Cualquier punto muy alejado de las funciones gaussianas podrían considerarse una anomalía. El conjunto de datos se dividió en dos subconjuntos de datos: conjunto de datos de entrenamiento y conjunto de datos de prueba. La prueba se realiza durante dos días de telemetría. El 20% final del conjunto de datos se utiliza para la prueba mientras que el otro 80% forma el conjunto de datos de entrenamiento. Un modelo es obtenido al



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

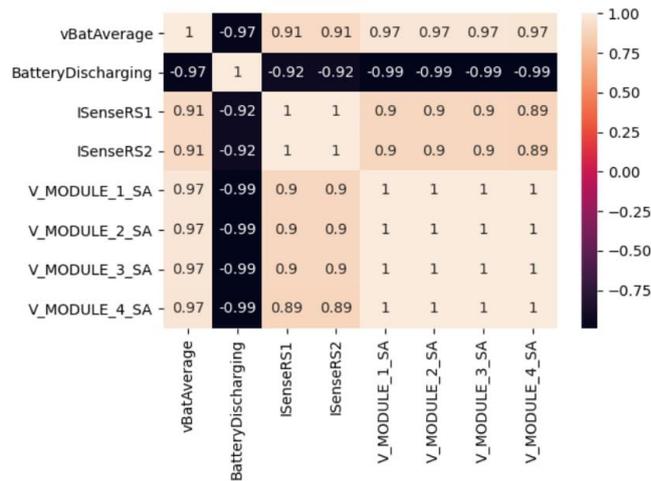


Fig. 1: Correlaciones entre valores de telemetría del subsistema de potencia

ejecutar el algoritmo sobre el conjunto de entrenamiento, donde la cantidad de componentes y el tipo de covarianza se seleccionan en un proceso iterativo que analiza information-theoretic criteria (BIC), cubriendo los 4 tipos de covarianza y la cantidad de componentes entre 1 y 20. La puntuación mínima de valores atípicos se establece como límite para la prueba.

2.2 Rolling Means

Rolling Means utiliza un enfoque estadístico simple para la detección de anomalías sobre un conjunto de datos de serie de tiempos. Dada una serie de referencias y una ventana de tamaño fijo N, el algoritmo obtiene primero la media de los N registros iniciales de la serie. Entonces la ventana se "mueve hacia adelante" en uno, recalculando la media de la ventana. Este proceso se repite hasta que la ventana final incluye el dato final. Una vez que todas las medias se ha obtenido, el algoritmo etiqueta como valores atípicos todos los puntos cuyo desvío de la media es S veces mayor que la desviación estándar que corresponde al punto.

Se aplica Rolling Means mediante un algoritmo [13] a cada subconjunto de datos, uno para cada variable de telemetría, y para cada uno genera un modelo de normalidad. Para utilizar este algoritmo, se debe establecer el tamaño de la ventana y un número fijo de desviaciones estándar. Para decidir el valor de estos parámetros, se ejecutan varias iteraciones con diferentes valores en el mismo conjunto de datos, y finalmente, conociendo la naturaleza de los datos y tomando el rol de experto, los mejores valores se utilizan.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Se eligió Rolling Means ya que es un algoritmo sensible a valores anómalos, siendo simple de implementar. Se basa en la desviación estándar, teniendo en cuenta el cambio en la serie de tiempo usando la ventana de tamaño fijo.

2.3 Otros métodos

También se probó el método de distribución normal multivariable, pero se descartó a favor de la Mezcla Gaussiana, dado que el primero necesita que sus datos sigan una distribución normal y no puede manejar varias campanas. Isolation Tree también fue analizado, pero se descartó dado que se etiquetaron incorrectamente la mayor parte de el conjunto de datos "normal" como anomalías, sin tener la posibilidad de utilizar un parámetro para cambiar su comportamiento.

3 Resultados

3.1 Modelo de Mezcla Gaussiana

Para obtener un grafico que nos brinde una aproximación visual al modelo generado utilizamos inicialmente solo dos características $V_MODULE_24_SA$ y $vBatAverage$ sobre los datos de entrenamiento. La figura 2 muestra graficamente las funciones gaussianas en verde y las diferentes agrupaciones generadas.

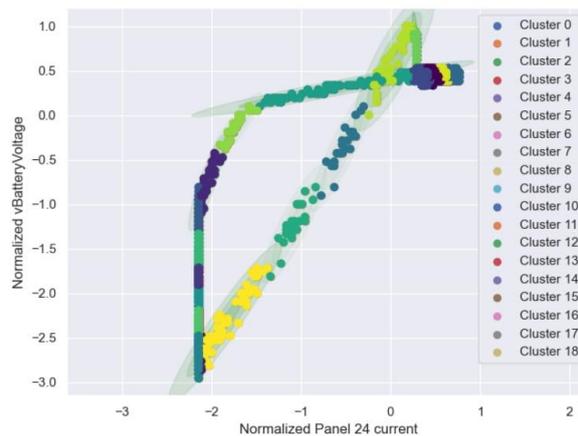


Fig. 2: Grupos o Clusters creados para 2 variables usando mezcla gaussiana



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

Se testean los datos de prueba con el modelo previamente generado. La figura 3 muestra como los datos de prueba sin modificación artificial ajustan al modelo.

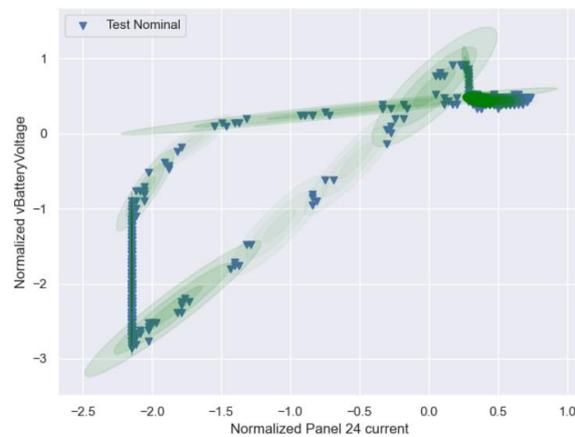


Fig. 3: Dataset de prueba sin anomalías generadas artificialmente. En verde las funciones gaussianas, todos los datos se ajustan al modelo. No hay falsos positivos

Los resultados con el conjunto de datos modificado artificialmente, con solo 2 variables ($V_MODULE_24_SA$ y $vBatAverage$), simulando corriente 0 en el panel solar 24, se muestran en la figura 4. Se detectan 880 anomalías. Si bien la caída progresiva de corriente no permite separar de forma clara cual dato es anómalo y cual no, la cantidad obtenida sobre el total de datos es una buena medida del estado general del sistema.

Si utilizamos las 28 características disponibles en el dataset 1 obtenemos 11 falsos positivos con el conjunto de datos sin modificar, es el 0,25% del conjunto de datos de prueba, si, también usando todas las características 1 pero con el dataset modificado artificialmente obtenemos 925 anomalías. Los resultados del experimento con 2 y 28 características, para datos originales o modificados artificialmente son mostrados en la tabla 2.

3.2 Rolling Means

El conjunto de datos se divide en 7 subconjuntos de datos, uno para cada variable de telemetría. Para cada uno de estos subconjuntos de datos, el algoritmo Rolling



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

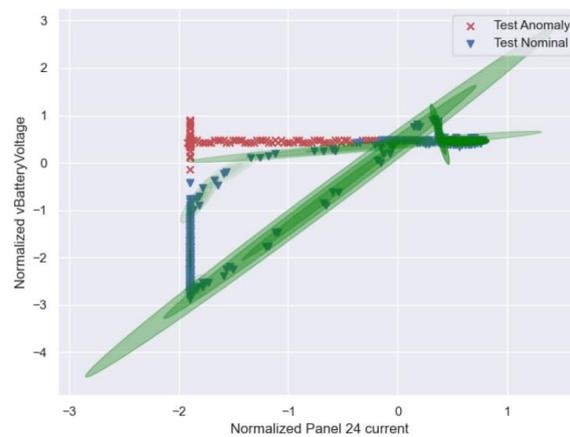


Fig. 4: Dataset de prueba con anomalías

# Variables	Dataset Normal	Dataset c/anomalías
2	0	880
28	11	925

Table 2: Anomalías detectadas por GMM

Means etiqueta los datos de cada variable como anomalías o no, según el "modelo de normalidad".

Ejecutando el algoritmo, con un tamaño de ventana de 1000 (una ventana que es la mitad del número de anomalías insertadas), utilizando 1 y 2 desviaciones estándar obtiene los próximos resultados. Cada subconjunto de datos se traza con líneas azules que representan datos considerados nominales y líneas rojas que representan los puntos de anomalías que detectó el algoritmo. La tabla 3 muestra, para una y dos desviaciones estándar la cantidad de anomalías detectadas en el dataset original y el modificado artificialmente.

4 Conclusiones

En el caso de Rolling Means, el modelo no es sensible a la correlación y define si un dato es anómalo basándose en la tendencia del conjunto de datos de series de tiempo. Un pico aislado en el gráfico se etiquetará como una anomalía, pero



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

#desviaciones estándar	Dataset Normal	Dataset c/anomalías
1	260	449
2	71	6

Table 3: Anomalías detectadas por Rolling Means para una y dos desviaciones estándar sobre dataset original y modificado

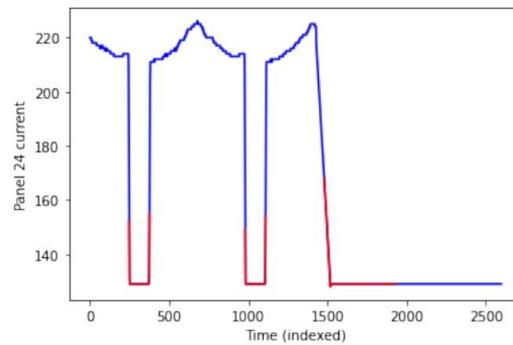


Fig. 5: Rolling Means aplicado al Panel 24 usando una desviación estándar

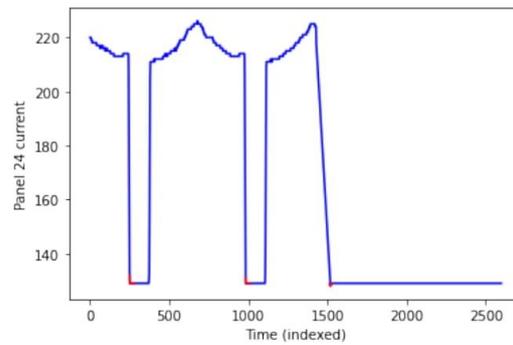


Fig. 6: Rolling Means aplicado al Panel 24 usando dos desviaciones estándar

puede ser el resultado de una acción contextual esperada. Aunque Rolling Means es un algoritmo simple, con pocas necesidades computacionales, al no tener en cuenta el contexto y las correlaciones no pueden manejar anomalías específicas y



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

dependientes del contexto. Usando una desviación estándar parece detectar las anomalías introducidas, pero también etiqueta erróneamente los datos válidos.

Usando dos desviaciones estándar, contrariamente a lo esperado, se comporta de la misma manera, etiquetando incorrectamente aún más datos. Rolling Means es un método válido para detectar valores atípicos producidos por ruido, pero no puede considerarse un algoritmo válido para la detección de anomalías. También necesita la intervención de un experto para establecer los parámetros iniciales. Por otro lado, el método de Mezcla Gaussiana muestra resultados prometedores. Se detectaron anomalías, sin etiquetar incorrectamente una gran cantidad de registros (signo de que el modelo no se ha sobreentrenado). Estas anomalías introducidas no pueden ser detectadas por los sistemas de control de límites dado que los valores probados son normales en un contexto determinado. La covarianza y la cantidad de clusters se obtuvieron automáticamente, sin un intervención experta.

5 Trabajo Futuro

Los resultados dan una vista informativa de los diferentes algoritmos, pero no pueden ser evaluado objetivamente ya que no hay datos etiquetados disponibles para compararlos. Si se pudieran obtener datos preetiquetados, se utilizarían métricas estadísticas para evaluar los resultados y ajustar los parámetros de los modelos para minimizar el número de falsos positivos producidos por el prototipo.

Entre los requerimientos típicos de estos sistemas se encuentra la detección de anomalías. Otros algoritmos como DBScan, y técnicas de aprendizaje profundo deben ser exploradas a futuro, la correlación de variables ha demostrado ser un capital valioso para detectar comportamientos anormales y pueden ofrecer una solución superadora al simple control de límites. Lamentablemente no es posible a la fecha encontrar un conjunto de datos satelital real con anomalías documentadas o fallos futuros que podrían detectarse mediante un análisis histórico de datos. Esta limitación es una de las mayores barreras a vencer en trabajos futuros si se pretende alcanzar en este área el mismo nivel de implementación que en otros sectores (Fraude Bancario, Intrusiones, Diagnósticos Médicos, etc).

References

1. Pablo Soligo and Jorge Salvador Ierache. Software de segmento terreno de próxima generación. In *XXIV Congreso Argentino de Ciencias de la Computación (La Plata, 2018)*, 2018.
2. Pablo Soligo and Jorge Salvador Ierache. Segmento terreno para misiones espaciales de próxima generación. *WICC 2019*.
3. Pablo Soligo, Jorge Salvador Ierache, and German Merkel. Telemetría de altas prestaciones sobre base de datos de serie de tiempos. 2020.
4. Unlam Ground Segment: Home unlam ground segment: Home. <https://ugs.unlam.edu.ar/>. Accessed: 2021-07-30.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019

5. Satnogs satnogs. <https://satnogs.org/>. Accessed: 2021-07-30.
6. Takehisa Yairi, Minoru Nakatsugawa, Koichi Hori, Shinichi Nakasuka, Kazuo Machida, and Naoki Ishihama. Adaptive limit checking for spacecraft telemetry data using regression tree learning. In *2004 IEEE International Conference on Systems, Man and Cybernetics (IEEE Cat. No. 04CH37583)*, volume 6, pages 5130–5135, IEEE, 2004.
7. Pablo Soligo and Jorge Salvador Ierache. Arquitectura de segmento terreno satelital adaptada para el control de límites de telemetría dinámicos. 2019.
8. Charu Aggarwal. *An introduction to outlier analysis*. Springer New York, 1 edition, 2017.
9. Aaron Rosenbaum. Detecting credit card fraud with machine learning. 2019.
10. Trevor Hastie, Robert Tibshirani, and Jerome Friedman. *The elements of Statistical Learning*. Springer Series in Statistics. Springer, 2 edition, 2008.
11. Low Orbit Satellite Dataset: Home low orbit satellite dataset: Home. <https://gidsa.unlam.edu.ar/data/LowOrbitSatellite.csv>. Accessed: 2021-07-30.
12. F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and E. Duchesnay. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, 12:2825–2830, 2011.
13. Algoritmo Rolling Means rollingmeans. <https://gidsa.unlam.edu.ar/data/rolling.py>. Accessed: 2021-07-30.



Código	FPI-009
Objeto	Guía de elaboración de Informe final de proyecto
Usuario	Director de proyecto de investigación
Autor	Secretaría de Ciencia y Tecnología de la UNLaM
Versión	5
Vigencia	03/9/2019



UNLaM - SECyT

FORMULARIO DE EVALUACIÓN DE ALUMNOS INTEGRANTES DE EQUIPOS DE INVESTIGACIÓN

FPI-013

Unidad Académica: Departamento Ingeniería e Investigaciones Tecnológicas (DIIT)
 Código: C230
 Título del Proyecto: Aprendizaje automático para el control del estado de salud en sistemas
 Aeroespaciales
 Director del Proyecto: Jorge Salvador Ierache
 Programa de acreditación: PROVINCE.. X., CyTMA2:.....
 Fecha de inicio:01/01/2020
 Fecha de finalización:31/12/2021

1. Datos del alumno

Apellido y Nombre: German Merkel
 DNI: 40853806
 Unidad Académica:DIIT
 Carrera que cursa:Ingeniería Informática
 Período evaluado: 01/01/2021 – 31/12/2021

2. Dictamen de evaluación de desempeño del alumno:

Colocar una cruz donde corresponda

- 2.1 Satisfactorio X
- 2.1 No satisfactorio:

Fundamentos del dictamen:

El alumno colaboró activamente en las modificaciones y adaptaciones para la publicación del UGS (<https://ugs.unlam.edu.ar/>). El alumno participó activamente en los trabajos presentados en WICC 2021 y propuestos para CACIC 2021 cumpliendo los objetivos planteados.

.....

.....

3. Propuesta de continuidad en el proyecto (si corresponde según duración estimada)

Colocar una cruz donde corresponda

- 3.1 Continuar en el presente proyecto X
- 3.2 No continuar en el presente proyecto:

Fundamentos del dictamen:

Su continuidad en el marco de la duración del proyecto que finaliza el 31/12/2021, se evalúa continuidad sujeta a disponibilidad del alumno en proyectos futuros

San Justo 26 agosto 2021

Lugar y fecha

Firma del Director

Jorge Salvador Ierache

Aclaración de firma