

CARRERA DEL INVESTIGADOR CIENTÍFICO Y TECNOLÓGICO

Informe Científico¹

PERIODO ²: 2017

1. DATOS PERSONALES

APELLIDO: Hasperué

NOMBRES: Waldo

Dirección Particular: Calle:

Localidad: Punta Lara

Dirección electrónica (donde desea recibir información):

2. TEMA DE INVESTIGACION

Aplicaciones en Big data. Tratamiento de flujos de datos

PALABRAS CLAVE (HASTA 3) Big data Stream processing Stream mining

3. DATOS RELATIVOS A INGRESO Y PROMOCIONES EN LA CARRERA

INGRESO: Categoría: asistente Fecha: 11/2016

ACTUAL: Categoría: asistente desde fecha: 11/2016

4. INSTITUCION DONDE DESARROLLA LA TAREA

Universidad y/o Centro: Universidad Nacional de La Plata

Facultad: Facultad de Informática

Departamento:

Cátedra:

Otros: Instituto de Investigación en Informática III-LIDI

Dirección: Calle: 50 y 120 N°: sn

Localidad: La Plata CP: 1900 Tel: (221) 4227707

Cargo que ocupa: Profesor adjunto dedicación semi-exclusiva

5. DIRECTOR DE TRABAJOS (En el caso que corresponda)

Apellido y Nombres: Lanzarini, Laura Cristina

Dirección Particular: Calle:

Localidad: La Plata CP:

Dirección electrónica:

.....
Firma del Director (si corresponde)

.....
Firma del Investigador

¹ Art. 11; Inc. "e"; Ley 9688 (Carrera del Investigador Científico y Tecnológico).

² El informe deberá referenciar a años calendarios completos. Ej.: en el año 2017 deberá informar sobre la actividad del período 1°-01-2015 al 31-12-2016, para las presentaciones bianuales. Para las presentaciones anuales será el año calendario anterior.

6. RESUMEN DE LA LABOR QUE DESARROLLA

Descripción para el repositorio institucional. Máximo 150 palabras.

Soy docente investigador en III-LIDI, instituto de la Facultad de Informática, UNLP e investigador del proyecto acreditado 11/F025 "Sistemas inteligentes. Aplicaciones en reconocimiento de patrones, minería de datos y big data" perteneciente al Programa de Incentivos, evaluado por la UNLP y financiado por el Ministerio de Educación (01/01/2018 – 31/12/2021). Entre las materias que dicto se encuentra una optativa llamada "Conceptos y aplicaciones en Big Data" de grado y de postgrado, cuyos contenidos están estrechamente vinculados a mi campo de investigación.

Mi área de estudio está enmarcada en el procesamiento de grandes volúmenes de datos ("Big Data") y en el estudio y desarrollo de técnicas adaptativas que permitan la extracción de conocimiento. En particular, técnicas y algoritmos que trabajen sobre flujos de datos ("Stream processing") mejorando la velocidad de respuesta, almacenando solo los datos más representativos y detectando las características principales de los datos, entre otras.

7. EXPOSICION SINTETICA DE LA LABOR DESARROLLADA EN EL PERIODO.

Debe exponerse, en no más de una página, la orientación impuesta a los trabajos, técnicas y métodos empleados, principales resultados obtenidos y dificultades encontradas en el plano científico y material. Si corresponde, explicita la importancia de sus trabajos con relación a los intereses de la Provincia.

Durante el período se han realizando diversos aportes usando el framework distribuido Spark Streaming.

En esta dirección, se comenzó con la dirección de un becario EVC quien se encuentra está investigando en una técnica de clustering dinámico basado en densidad que se ejecuta de manera distribuida.

En la misma línea se ha trabajado en el procesamiento en streaming de datos de transacciones de criptomonedas. Desarrollando para ello estrategias que aplican técnicas de machine learning que presentan la característica de ser iterativas, operando sobre el conjunto completo de los datos de un flujo, brindando resultados en tiempos de respuestas cortos los cuales se adaptan de manera dinámica a la llegada de nuevos datos. Los resultados de estos ensayos han sido utilizados en un flujo de transacciones de criptomonedas, cuyo análisis pudo ser publicado en una revista.

Se ha implementado en el framework Spark Streaming una aplicación que permite el cálculo del índice de Hertz de manera online y dinámica, esto es, cada cierto tiempo la aplicación usa los nuevos datos recolectados y los procesa con aquellos que habían sido procesados previamente para poder hacer un seguimiento online de un cierto mercado de negocios. Los resultados de esta investigación fueron publicados en un congreso.

En el área del procesamiento de texto y el análisis de sentimientos se logró finalizar una tesis de grado, al mismo tiempo que se comenzó con otras tres.

8. TRABAJOS DE INVESTIGACION REALIZADOS O PUBLICADOS EN ESTE PERIODO.

8.1 PUBLICACIONES. *Debe hacer referencia exclusivamente a aquellas publicaciones en las que haya hecho explícita mención de su calidad de Investigador de la CIC (Ver instructivo para la publicación de trabajos, comunicaciones, tesis, etc.). Toda publicación donde no figure dicha mención no debe ser adjuntada porque no será tomada en consideración. A cada publicación, asignarle un número e indicar el nombre de los autores en el mismo orden que figuran en ella, lugar donde fue publicada, volumen, página y año. A continuación, transcribir el resumen (abstract) tal como aparece en la publicación. La copia en papel de cada publicación se presentará por separado. Para cada publicación, el investigador deberá, además, aclarar el tipo o grado de participación que le cupo en el desarrollo del trabajo y, para aquellas en las que considere que ha hecho una contribución de importancia, deberá escribir una breve justificación. Asimismo, para cada publicación deberá indicar si se encuentra depositada en el repositorio institucional CIC-Digital.*

En el período informado se publicaron tres trabajos. En lo que se refiere al primer artículo cuyo título es "Some stylized facts of the bitcoin market", publicado en octubre de 2017, deseo aclarar que la omisión en mi filiación como Investigador Asociado CIC se debe a que fue enviado para su revisión en noviembre de 2016 período que coincide con mi vinculación a la CIC. La aceptación de este trabajo nos fue notificada en abril de 2017 no habiendo luego oportunidad de introducir modificaciones.

1. Fernandez Bariviera, Aurelio; Basgall, María José; Hasperué, Waldo; Naiouf, Marcelo. Some stylized facts of the bitcoin market. Physica A - Statistical and theoretical physics. Elsevier science bv. Vol.484 n°. P82 - 90. ISSN 0378-4371. 2017.

Abstract. In recent years a new type of tradable assets appeared, generically known as cryptocurrencies. Among them, the most widespread is Bitcoin. Given its novelty, this paper investigates some statistical properties of the Bitcoin market. This study compares Bitcoin and standard currencies dynamics and focuses on the analysis of returns at different time scales. We test the presence of long memory in return time series from 2011 to 2017, using transaction data from one Bitcoin platform. We compute the Hurst exponent by means of the Detrended Fluctuation Analysis method, using a sliding window in order to measure long range dependence. We detect that Hurst exponents changes significantly during the first years of existence of Bitcoin, tending to stabilize in recent times. Additionally, multiscale analysis shows a similar behavior of the Hurst exponent, implying a self-similar process.

Tipo de participación: preparación de los datos, implementación de algoritmos para el análisis de los mismos y escritura de los aspectos técnicos del paper.

No se encuentra en el repositorio institucional CIC-Digital.

2. Basgall, María José; Hasperué, Waldo; Naiouf, Marcelo; Fernandez Bariviera, Aurelio. Cálculo del exponente de Hurst utilizando Spark Streaming: enfoque experimental sobre un flujo de transacciones de criptomonedas. XXIII Congreso Argentino de Computación. La Plata. Argentina. 2017.

Resumen. En la actualidad es cada vez más común encontrarse con problemas de Big Data, donde las aplicaciones para tratar estos datos son implementadas en frameworks específicos. Uno de ellos, cada vez es más utilizado, es Apache Spark. Uno de los modos de trabajo de Apache Spark, se obtiene mediante el módulo Spark Streaming que permite el tratamiento de datos provenientes de un flujo de información potencialmente infinito. En este trabajo presentamos una aplicación implementada en Spark Streaming que realiza el cálculo del exponente de Hurst, un indicador muy utilizado en el análisis de mercado para la detección de memoria a largo plazo. Los ensayos realizados se hicieron sobre flujos simulados de transacciones de criptomonedas que demuestran la capacidad de Spark Streaming para el tratamiento de este tipo de flujos.

Tipo de participación: preparación de los datos, implementación de algoritmos para el análisis de los mismos y escritura de los aspectos técnicos del paper.

No se encuentra en el repositorio institucional CIC-Digital.

3. Lanzarini, Laura; Hasperué, Waldo; Estrebow, César; Ronchetti, Franco; Villa Monte, Augusto; Aquino, Germán; Quiroga, Facundo; Basgall, María José; Rojas, Luis; Corvi, Julieta; Jimbo Santana, Patricia; Fernandez Bariviera, Aurelio; Puente, Cristina; Olivas Varela, José Angel. Minería de datos y big data. Aplicaciones en

señales y textos. Workshop de investigadores en ciencia de la computación. ITBA. Ciudad autónoma de Buenos Aires. Argentina. 2017.

Resumen. Esta línea de investigación se centra en el estudio y desarrollo de Sistemas Inteligentes para la resolución de problemas de Minería de Datos y Big Data utilizando técnicas de Aprendizaje Automático. Los sistemas desarrollados se aplican particularmente al procesamiento de señales y textos.

Con respecto al Reconocimiento de Señales el énfasis está puesto en el análisis de videos con el objetivo de identificar acciones humanas que faciliten la interfaz hombre/máquina y en la detección de patrones de movimiento en videos provenientes de cámaras de vigilancia asociados con situaciones de riesgo.

En el área de la Minería de Datos se está trabajando, por un lado, en la generación de un modelo de fácil interpretación a partir de la extracción de reglas de clasificación que permita justificar la toma de decisiones y, por otro lado, en el desarrollo de nuevas estrategias para tratar grandes volúmenes de datos.

Con respecto a Minería de Textos se han desarrollado métodos capaces de extraer las palabras clave de documentos independientemente del lenguaje. Además, se han desarrollado estrategias para resumir documentos a través de la extracción utilizando métricas de selección y técnicas de optimización de los párrafos más representativos.

Tipo de participación: Escritura del 20% del trabajo

No se encuentra en el repositorio institucional CIC-Digital.

8.2 TRABAJOS EN PRENSA Y/O ACEPTADOS PARA SU PUBLICACIÓN. *Debe hacer referencia exclusivamente a aquellos trabajos en los que haya hecho explícita mención de su calidad de Investigador de la CIC (Ver instructivo para la publicación de trabajos, comunicaciones, tesis, etc.). Todo trabajo donde no figure dicha mención no debe ser adjuntado porque no será tomado en consideración. A cada trabajo, asignarle un número e indicar el nombre de los autores en el mismo orden en que figurarán en la publicación y el lugar donde será publicado. A continuación, transcribir el resumen (abstract) tal como aparecerá en la publicación. La versión completa de cada trabajo se presentará en papel, por separado, juntamente con la constancia de aceptación. En cada trabajo, el investigador deberá aclarar el tipo o grado de participación que le cupo en el desarrollo del mismo y, para aquellos en los que considere que ha hecho una contribución de importancia, deberá escribir una breve justificación.*

8.3 TRABAJOS ENVIADOS Y AUN NO ACEPTADOS PARA SU PUBLICACION. *Incluir un resumen de no más de 200 palabras de cada trabajo, indicando el lugar al que han sido enviados. Adjuntar copia de los manuscritos.*

8.4 TRABAJOS TERMINADOS Y AUN NO ENVIADOS PARA SU PUBLICACION. *Incluir un resumen de no más de 200 palabras de cada trabajo.*

8.5 COMUNICACIONES. *Incluir únicamente un listado y acompañar copia en papel de cada una. (No consignar los trabajos anotados en los subtítulos anteriores).*

8.6 INFORMES Y MEMORIAS TECNICAS. *Incluir un listado y acompañar copia en papel de cada uno o referencia de la labor y del lugar de consulta cuando corresponda. Indicar en cada caso si se encuentra depositado en el repositorio institucional CIC-Digital.*

9. TRABAJOS DE DESARROLLO DE TECNOLOGÍAS.

9.1 DESARROLLOS TECNOLÓGICOS. *Describir la naturaleza de la innovación o mejora alcanzada, si se trata de una innovación a nivel regional, nacional o internacional, con qué financiamiento se ha realizado, su utilización potencial o actual por parte de empresas u otras entidades, incidencia en el mercado y niveles de facturación del respectivo producto o servicio y toda otra información conducente a demostrar la relevancia de la tecnología desarrollada.*

9.2 PATENTES O EQUIVALENTES *Indicar los datos del registro, si han sido vendidos o licenciados los derechos y todo otro dato que permita evaluar su relevancia.*

9.3 PROYECTOS POTENCIALMENTE TRANSFERIBLES, NO CONCLUIDOS Y QUE ESTAN EN DESARROLLO. *Describir objetivos perseguidos, breve reseña de la labor realizada y grado de avance. Detallar instituciones, empresas y/o organismos solicitantes.*

9.4 OTRAS ACTIVIDADES TECNOLÓGICAS CUYOS RESULTADOS NO SEAN PUBLICABLES *(desarrollo de equipamientos, montajes de laboratorios, etc.).*

9.5 *Sugiera nombres (e informe las direcciones) de las personas de la actividad privada y/o pública que conocen su trabajo y que pueden opinar sobre la relevancia y el impacto económico y/o social de la/s tecnología/s desarrollada/s.*

10. SERVICIOS TECNOLÓGICOS. *Indicar qué tipo de servicios ha realizado, el grado de complejidad de los mismos, qué porcentaje aproximado de su tiempo le demandan y los montos de facturación.*

11. PUBLICACIONES Y DESARROLLOS EN:

11.1 DOCENCIA

11.2 DIVULGACIÓN

En cada caso indicar si se encuentran depositados en el repositorio institucional CIC-Digital.

12. DIRECCION DE BECARIOS Y/O INVESTIGADORES. *Indicar nombres de los dirigidos, Instituciones de dependencia, temas de investigación y períodos.*

1. Beca CONICET. Becaria María José Basgall. a) Big Data sobre redes sociales aplicado a situaciones de emergencias humanitarias utilizando Cómputo de Altas Prestaciones Abril 2016 - Marzo 2021

2. Beca CIN-EVC. Becario Roberto Molina. Implementación de una técnica de aprendizaje automático para el tratamiento de flujos de datos. Abril 2017 - Marzo 2018

13. DIRECCION DE TESIS. *Indicar nombres de los dirigidos y temas desarrollados y aclarar si las tesis son de maestría o de doctorado y si están en ejecución o han sido defendidas; en este último caso citar fecha.*

1. Tesina de grado de Juan Agustín Tibaldo e Ignacio Saporiti. Defendida en septiembre de 2017

2. Tesina de grado de Juan Manuel Coria. En ejecución

3. Tesina de grado de Roberto Molina. En ejecución

4. Tesina de grado de Emiliano Gianneto y Lucía Saporitti. En ejecución.

- 14. PARTICIPACION EN REUNIONES CIENTIFICAS.** *Indicar la denominación, lugar y fecha de realización, tipo de participación que le cupo, títulos de los trabajos o comunicaciones presentadas y autores de los mismos.*
XXIII Congreso Argentino de Computación. La Plata. Argentina. Autor del trabajo "Cálculo del exponente de Hurst utilizando Spark Streaming: enfoque experimental sobre un flujo de transacciones de criptomonedas". Octubre de 2017.
- 15. CURSOS DE PERFECCIONAMIENTO, VIAJES DE ESTUDIO, ETC.** *Señalar características del curso o motivo del viaje, período, instituciones visitadas, etc.*
- 16. SUBSIDIOS RECIBIDOS EN EL PERIODO.** *Indicar institución otorgante, fines de los mismos y montos recibidos.*
1. Universidad Nacional de La Plata (Convocatoria 2016-2017). Subsidio para viajes y estadía (Ecuador 21/06/2017 - 01/07/2017). Monto: \$10500
2. Comisión de Investigaciones Científicas. Subsidio semi-automático para insumos. Monto: \$8000
- 17. OTRAS FUENTES DE FINANCIAMIENTO.** *Describir la naturaleza de los contratos con empresas y/o organismos públicos.*
- 18. DISTINCIONES O PREMIOS OBTENIDOS EN EL PERIODO.**
- 19. ACTUACION EN ORGANISMOS DE PLANEAMIENTO, PROMOCION O EJECUCION CIENTIFICA Y TECNOLÓGICA.** *Indicar las principales gestiones realizadas durante el período y porcentaje aproximado de su tiempo que ha utilizado.*
- 20. TAREAS DOCENTES DESARROLLADAS EN EL PERIODO.** *Indicar el porcentaje aproximado de su tiempo que le han demandado.*
Profesor adjunto con dedicación semi-exclusiva de las materias "Conceptos y Aplicaciones en Big Data" y "Taller de programación".
Dedicación total: 32 semanas al año, 9 horas por semana (30% del tiempo)
- 21. OTROS ELEMENTOS DE JUICIO NO CONTEMPLADOS EN LOS TITULOS ANTERIORES.** *Bajo este punto se indicará todo lo que se considere de interés para la evaluación de la tarea cumplida en el período.*
- 22. TITULO, PLAN DE TRABAJO A REALIZAR EN EL PROXIMO PERIODO.** *Desarrollar en no más de 3 páginas. Si corresponde, explicité la importancia de sus trabajos con relación a los intereses de la Provincia.*
Título: Aplicaciones en Big data. Tratamiento de flujos de datos

Motivación

El minado de flujos de datos (stream mining) es un paradigma que permite llevar a cabo tareas de recolectar, procesar y analizar grandes cantidades de datos que se generan de manera online (data stream), las cuales crecen a una tasa sin precedentes realizando dichas tareas de forma eficiente, escalable y en tiempo real. Como los datos conforman un flujo, potencialmente infinito, es necesario o bien procesarlos inmediatamente o bien almacenarlos, ya que de lo contrario se pierden para siempre.

Por otro lado, las técnicas de aprendizaje automático (machine learning) son muy utilizadas para conseguir modelos que consiguen un resumen de los datos pudiendo ofrecer conocimiento útil sobre los mismos.

La desventaja más grande que poseen las técnicas adaptativas de machine learning es que los tiempos de respuesta comienzan a no ser viables a medida que crece el volumen de información a analizar. Los algoritmos de clustering, clasificación, predicción o incluso los algoritmos de redes neuronales necesitan realizar varias iteraciones sobre el conjunto completo de datos para obtener un resultado útil. Esta característica iterativa, cuando los datos provienen de un flujo potencialmente infinito, es un gran problema a resolver [1] [2] [6] [8] [12] [15], ya que recolectado parte del flujo, al mismo tiempo que se lo procesa, se continúa con la recolección de más datos, los cuales hacen crecer de manera continua el conjunto de datos.

Para ello, es necesario contar con técnicas que mejoren su velocidad de respuesta, que almacenen solo aquellos datos que son representativos del flujo, que detecten las características principales de los datos del flujo, entre otras.

Estado del arte

Los flujos de datos son una fuente de datos muy interesante para su análisis, dada su naturaleza de fuente infinita y los desafíos que presenta en términos de eficiencia. Es un tema muy estudiado en la actualidad y diversos son los aportes que se han hecho al respecto. En el campo de la astronomía, se ha desarrollado un framework que permite al procesamiento con alto rendimiento de datos capturados por radio [5].

El área de investigación conocida como Internet de las Cosas (IoT) básicamente se compone de una red de sensores de diversos tipos que generan un flujo de datos de alta frecuencia. En este sentido, en [13] se propone un estudio para la detección de anomalías en este tipo de escenario, mientras que en [11] presentan un método híbrido de análisis de texto y machine learning para la detección de falsos positivos en sensores utilizados para sistemas de seguridad. En el trabajo de [7] se presenta el uso de un sistema denominado ELPACO que permite analizar de manera eficiente la gran cantidad de datos que se generan durante los experimentos en el CERN.

Por otro lado, ciertos autores presentan soluciones de alto nivel para la escritura fácil, rápida y correcta de aplicaciones para el tratamiento de flujos de datos. En [8] se proponen plantillas de alto nivel para el desarrollo de aplicaciones mediante el uso de módulos que resuelven problemas muy específicos que son comunes en el análisis de flujos de datos. En [9] presentan un nuevo lenguaje de programación el cual posee la característica de ser extensible ofreciendo así una completa personalización al programador.

[14] presenta una aplicación que consume varios flujos provenientes de diferentes dispositivos (IoT) los cuales son analizados y utilizados en varias aplicaciones, presentado en un nuevo paradigma emergente denominado Fog.

Objetivo

El presente plan de trabajo tiene su eje central en la investigación, implementación y aplicación de técnicas y algoritmos que sirvan para resolver aplicaciones de Big Data, en particular la extracción de conocimiento a partir de flujos de datos. Continuando así con la línea de investigación del proyecto en el cual se enmarca este plan de trabajo [1] [2] [3] [4].

Los flujos de datos que se pretenden analizar y que resulta de mayor interés son los generados por las redes sociales. Las técnicas desarrolladas para estos flujos permitirían llevar a cabo tareas de clustering o clasificación, las cuales pueden ser utilizadas para aplicaciones como el análisis de sentimiento. El análisis de flujos de mensajes de redes sociales, en particular Twitter, que es una red pública donde la cantidad de información es muy grande, implica el uso de arquitecturas que permitan la ejecución de algoritmos de manera paralela o distribuida. Por ello las técnicas estudiadas se implementan en nuevos frameworks de ambientes distribuidos como lo es Apache Spark.

Puntualmente se están implementando rastreadores web (crawler) que obtenga la información pública de diferentes redes sociales como pueden ser Twitter, Facebook o comentarios de noticias en portales web, entre otros, analizando el contenido con el fin de buscar comentarios afines a eventos que resulten de interés.

Otros proyectos estudiados que tienen a los flujos de datos como fuente de información son aquellos que provienen de una red de sensores. Estos flujos son generados por diferentes sensores como pueden ser, temperatura, luminosidad, movimiento, etc. que son recolectados de manera continua generando así un volumen de información muy grande. Este tipo de problemas tiene su principal aporte en detectar los cambios cambios que se sucedan en la distribución de los datos del flujo, cambios que en principio, serían producidos por una anomalía.

Un tercer tipo de flujo de dato estudiado son los generados por las transacciones de las criptomonedas en donde interesa hacer estudio del mercado y predicciones del mismo.

En particular es interesante poder tratar de manera masiva la enorme cantidad de datos generados, este plan de trabajo se encuentra enmarcado con las líneas de investigación actuales del grupo de trabajo en el III-LIDI (data streaming y big data).

Bibliografía

1. Basgall M. J. et al. Un Enfoque Dinámico para la Detección de Relaciones entre Tópicos en Textos provenientes de Redes Sociales. Proceedings de las III Jornadas de Cloud Computing y Big Data. Julio de 2015.
2. Basgall M. J., Hasperué W., Naiouf M. Tratamiento de un flujo de datos usando ventanas deslizantes con MapReduce. Proceedings de las IV Jornadas de Cloud Computing y Big Data. Julio de 2016.
3. Basgall, M. J.; Hasperué, W.; Naiouf, M.; Fernandez Bariviera, A. Cálculo del exponente de Hurst utilizando Spark Streaming: enfoque experimental sobre un flujo de transacciones de criptomonedas. XXIII Congreso Argentino de Ciencias de la Computación. La Plata, Argentina. Octubre de 2017.
4. Basgall, M.J., Hasperué, W., Naiouf, M., Fernández, A. & Herrera, F. SMOTE-BD: An Exact and Scalable Oversampling Method for Imbalanced Classification in Big Data. Proceedings de las VI Jornadas de Cloud Computing y Big Data. Julio de 2018.
5. Cranmer, M.D.; Barsdell, B.R.; Price, D.C.; Dowell, J; et al. Bifrost: A Python/C++ Framework for High-Throughput Stream Processing in Astronomy. Journal of Astronomical Instrumentation, Volume 6(4): 175-194. 2017.
6. Desai, D. and Joshi, A. (2015) A deviant load shedding system for data stream mining. Procedia Computer Science, 45:118-126, 2015. International Conference on Advanced Computing Technologies and Applications (ICACTA).
7. Lindemann, T; Kauke, J. and Teubner, J. Efficient Stream Processing of Scientific Data. IEEE 34th International Conference on Data Engineering Workshops (ICDEW), Paris, pp. 140-145. 2018. doi: 10.1109/ICDEW.2018.00029.
8. Mamouras, K., Stanford, C., Alur, R., Ives, Z.G., & Tannen, V. Data-Trace Types for Distributed Stream Processing Systems. 2017.
9. Martin Hirzel, Scott Schneider, and Buğra Gedik. SPL: An Extensible Language for Distributed Stream Processing. ACM Trans. Program. Lang. Syst. 39, 1, Article 5 (March 2017), 39 pages. DOI: <https://doi.org/10.1145/3039207>.
10. Molina, R. & Hasperué, W. D3CAS: un Algoritmo de Clustering para el Procesamiento de Flujos de Datos en Spark. XXIII Congreso Argentino de Ciencias de la Computación. La Plata, Argentina. Octubre de 2018.
11. Sima, A.-C.; Stockinger, K.; Affolter, K.; Braschler, M.; Monte, P. and Kaiser, L. A hybrid approach for alarm verification using stream processing, machine learning and text analytics. Proceedings of the 21st International Conference on Extending Database Technology. March 26-29, 2018.
12. Stefanowski, J., Cuzzocrea, A., and Slezak, D. (2014). Processing and mining complex data streams. Inf. Sci., 285:63-65.
13. Tatbul, N. New Challenges and Opportunities in Stream Processing: Transactions, Predictive Analytics, and Beyond. In Proceedings of the 12th ACM International

Conference on Distributed and Event-based Systems (DEBS '18). ACM, New York, NY, USA, 14-15. 2018. DOI: <https://doi.org/10.1145/3210284.3214706>.

14. Yang, S. IoT Stream Processing and Analytics in The Fog. CoRR, Volume abs/1705.05988. 2017.

15. Zare Moodi, P., Beigy, H., and Kamali Siahroudi, S. (2015) Novel class detection in data streams using local patterns and neighborhood graph. Neurocomput., 158(C):234-245, June 2015.

Condiciones de la presentación:

- A. El Informe Científico deberá presentarse dentro de una carpeta, con la documentación abrochada y en cuyo rótulo figure el Apellido y Nombre del Investigador, la que deberá incluir:
- Una copia en papel A-4 (puntos 1 al 22).
 - Las copias de publicaciones y toda otra documentación respaldatoria, en otra carpeta o caja, en cuyo rótulo se consignará el apellido y nombres del investigador y la leyenda "Informe Científico Período".
 - Informe del Director de tareas (en los casos que corresponda), en sobre cerrado.
- B. Envío por correo electrónico:
- Se deberá remitir por correo electrónico a la siguiente dirección: infinvest@cic.gba.gob.ar (puntos 1 al 22), en formato .doc zipeado, configurado para papel A-4 y libre de virus.
 - En el mismo correo electrónico referido en el punto a), se deberá incluir como un segundo documento un currículum resumido (no más de dos páginas A4), consignando apellido y nombres, disciplina de investigación, trabajos publicados en el período informado (con las direcciones de Internet de las respectivas revistas) y un resumen del proyecto de investigación en no más de 250 palabras, incluyendo palabras clave.
- C. Sistema SIBIPA:
- Se deberá peticionar el informe en la modalidad on line, desde el sitio web de la CIC, sistema SIBIPA (ver instructivo).

Nota: El Investigador que desee ser considerado a los fines de una promoción, deberá solicitarlo en el formulario correspondiente, en los períodos que se establezcan en los cronogramas anuales.