

UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE QUITO

CARRERA:
INGENIERÍA DE SISTEMAS

Trabajo de Titulación previo a la obtención del título de:
Ingeniero de Sistemas

TEMA:
**ESTADO DEL ARTE DE LA TENDENCIA DE LA CIENCIA DE DATOS
APLICADA A LA ADMINISTRACIÓN DE REDES DE COMPUTADORES**

AUTOR:
CLAUDIO VINICIO AGUIRRE LOZANO

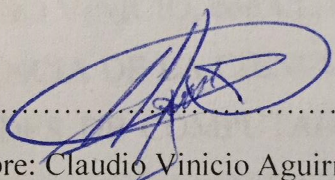
TUTOR:
JOSÉ LUIS AGUAYO MORALES

Quito, agosto de 2018

CESIÓN DE DERECHOS DE AUTOR

Yo, Claudio Vinicio Aguirre Lozano, con documento de identificación N° 1715061089, manifiesto mi voluntad y cedo a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que soy autor del trabajo de titulación intitulado: “ESTADO DEL ARTE DE LA TENDENCIA DE LA CIENCIA DE DATOS APLICADA A LA ADMINISTRACIÓN DE REDES DE COMPUTADORES”, mismo que ha sido desarrollado para optar por el título de: INGENIERO DE SISTEMAS, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En aplicación a lo determinado en la Ley de Propiedad Intelectual, en mi condición de autor me reservo los derechos morales de la obra antes citada. En concordancia, suscribo este documento en el momento que hago entrega del trabajo final en formato impreso y digital a la Biblioteca de la Universidad Politécnica Salesiana.



Nombre: Claudio Vinicio Aguirre Lozano
Cédula: 1715061089

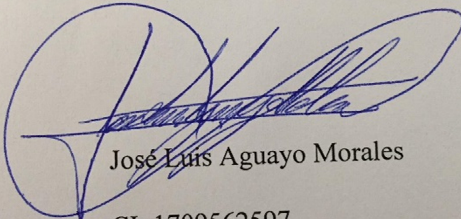
Quito, agosto del 2018

DECLARATORIA DE COAUTORÍA DEL TUTOR

Yo declaro que bajo mi dirección y asesoría fue desarrollado el Artículo Académico, con el tema: ESTADO DEL ARTE DE LA TENDENCIA DE LA CIENCIA DE DATOS APLICADA A LA ADMINISTRACIÓN DE REDES DE COMPUTADORES realizado por CLAUDIO VINICIO AGUIRRE LOZANO, obteniendo un producto que cumple con todos los requisitos estipulados por la Universidad Politécnica Salesiana, para ser considerados como trabajo final de titulación.

Quito, agosto del 2018

Atentamente,



José Luis Aguayo Morales
CI: 1709562597

ESTADO DEL ARTE DE LA TENDENCIA DE LA CIENCIA DE DATOS APLICADA A LA ADMINISTRACIÓN DE REDES DE COMPUTADORES

Aguayo Morales José Luis¹, Aguirre Lozano Claudio Vinicio²
Universidad Politécnica Salesiana
CARRERA DE INGENIERÍA DE SISTEMAS
Quito

Resumen-

Este artículo determinará el estado del arte de la ciencia de datos aplicada a redes computacionales.

Para el análisis se toman en cuenta los estudios y aplicaciones realizadas, casos de éxito o fracaso, los lineamientos teóricos generales, directrices y herramientas particulares del análisis. El análisis de la ciencia de los datos permite obtener una estrategia para tomar decisiones más acertadas, en base a tendencias en el uso de la infraestructura computacional.

La metodología científica es el procedimiento investigativo utilizado esencialmente en la creación de conocimiento basado en las ciencias, el mismo que se ha utilizado en este artículo.

El resultado final indica que existe un incremento en la toma de datos en los nodos que conforman los data centers, tanto en sus componentes internos como en las interconexiones. Cabe destacar que también ya existen diferentes aplicaciones comerciales que permiten optimizar, centralizar la información para los diferentes administradores de estas plataformas, y proyectar su comportamiento, utilizando técnicas de ciencias de datos, para el cálculo de grandes cantidades de variables.

Se encuentra que el crecimiento de datos a analizarse es relevante, la adquisición de nuevos datos permite encontrar nuevas relaciones causales que anteriormente no se consideraban, por ejemplo, la temperatura del CPU no era considerada como factor determinante para la administración, pero ahora este es un factor clave que suma al análisis de consumo energético en ambientes corporativos.

Abstract-

This article will show the state of art about data science applied to computational networks.

For this work, the analysis was realized by the studies and the applications related, besides, success and fail study cases, theories, guide lines, tools, and so on used to data science.

The research question is, if the data science analysis can be used to help to better decisions, by of trending in technological infrastructure that showed in their output data.

The scientific methodology is the procedure investigative used essentially in the creation of knowledge based on the science, the same that has been used in this article

The results finded showed an increase of measures of output data, especially in nodes of data centers in their internal measures and interactions each other's. Also, worth highlighting is business applications existed to optimize and centralize information for a better manage by techniques of data science when a millions of data must be evaluated.

Finally, one of the results is the growing of data is significant, because by the analysis of many data showed new relationships that before were not be considered, by example: the CPU temperature and data management to got a better efficiently of power.

¹ Ingeniero en Electrónica Y Telecomunicaciones, Docente de la carrera de Ingeniera de Sistemas – Universidad Politécnica Salesiana, Sede Quito – Campus Sur. Autor para correspondencia: jaguayo@ups.edu.ec

² Estudiante de Ingeniería de Sistemas-Universidad Politécnica Salesiana, Sede Quito – Campus Sur, Autor para correspondencia: caguirrelo@est.ups.edu.ec; aguirreclaudio@hotmail.com

I. INTRODUCCIÓN

La explosión de redes informáticas por la necesidad de los negocios, empresas y corporaciones, entre otros, de tener una comunicación directa entre sus empleados, oficinas, proveedores, etc., forma los departamentos de IT, quienes se encargan del control de los activos que se interconectan en las redes, surge de esta manera la necesidad de tener el control y el conocimiento de los diferentes cambios, problemas, inconvenientes que se pueden presentar en el correcto funcionamiento de las comunicaciones.

La necesidad de conocer cómo se puede extraer información de una red de computadores desde la infinidad de datos que generan las mismas, surgen al analizar parámetros como: velocidad, costo, seguridad, disponibilidad, escalabilidad, confiabilidad, topología, entre otros [1], [2].

Es necesario entonces responder a inquietudes como: ¿Qué técnicas existen? [3] ¿Cuáles se pueden aplicar para el manejo de los datos que reportan las redes de computadoras? [3] [4] ¿Cómo se puede obtener información para predecir los acontecimientos desde una inmensa cantidad de datos? [4] ¿Es una solución la ciencia de datos? [3], [4] ¿Cuál es la tendencia de la ciencia de datos aplicada a la administración de redes? [3].

La tendencia encontrada en la investigación, muestra que cada día se recolecta cantidad de datos, que son analizados con diferentes técnicas, permitiendo a los departamentos de TI, tener la capacidad de mantener un monitoreo constante de su infraestructura y prevenir con anticipación situaciones que puedan inutilizada la misma.

II. ANÁLISIS DE LA INFORMACIÓN

Partiendo del concepto de dato dado por [5] como la representación simbólica de números o letras que han sido recopilados como información la misma que puede ser cualitativa o cuantitativa, facilitando la deducción del análisis que se requiera, los datos muestran escenarios que por si mismos, no contribuyen con información importante, y toma valor en contexto, ligado a la experiencia y al análisis para así adquirir un valor relevante.

La técnica analítica o de análisis de datos según [5] toma los datos, los filtra, los analiza y se obtiene la información útil para la toma de

decisiones, prevención de adversidades y determinación de tendencias. Instrumentos de análisis de datos permiten, entre otras ampliar la visión de los negocios, de optimizar la producción de bienes y servicios como referencia [6], es decir el análisis de datos potencia la manera en la cual se examinan los productos, servicios, clientes, mercados, o tendencias, para desarrollar soluciones enfocadas en atender las prioridades y necesidades de cada industria y empresa referencia. [5].

De acuerdo a lo revisado [7], el análisis de datos determina comportamientos futuros basados en algoritmos inteligentes, que evolucionan con el tiempo por el constante flujo de nueva información que enriquecen los patrones de análisis.

Los datos se clasifican de la siguiente manera: estructurados, desestructurados y semi estructurados.

A. Clasificación de los datos

1) *Datos estructurados*: Principalmente es información que se encuentra almacenada en bases de datos, conformados claramente por filas y columnas que se pueden ordenar y clasificar de una manera sencilla, donde todo esta ordenado, clasificado; 2) *Datos desestructurados*: Los datos no estructurados, generalmente son datos binarios, que no poseen una estructura, es una gran asociación de objetos desorganizados, datos como el correo electrónico, archivos de texto, hojas de cálculo Excel, entre otros, [8]; 3) *Datos semiestructurados*: Los datos semi estructurados son una mezcla de los dos descritos anteriormente, no presentan una clara estructura, pero si presentan una organización definida en la metadata, donde se observa relaciones con los objetos, claramente aceptados como los formatos estandarizados de XML, Json, entre otros.

El proceso de analizar los datos permite interpretar la información que ha sido recolectada con el objetivo de identificar el peso que se puede tener, es por este motivo que estos tipos de análisis se clasifican en análisis cualitativo y análisis cuantitativo.

El análisis cualitativo recolecta datos sin medición numérica con el fin de construir teorías, descubrir o afinar preguntas, revelar situaciones, describir sucesos, hechos o patrones y explicarlos para que ayuden en la toma de decisiones [8], estos están basados en la observación, las escalas de actitudes, el análisis de contenido o los cuestionarios, y sus instrumentos lo hacen con la

observación, las historias de vida, las entrevistas, los grupos de enfoques o los documentos.

El análisis cuantitativo busca probar hipótesis y teorías, o establecer patrones de comportamiento a partir de una recolección de datos del mismo tipo, con base a la medición numérica y análisis estadístico. Para ello se utilizan instrumentos de recolección de datos cuantitativos [8]. Los cuantitativos pueden dividirse en: empíricos analíticos (descriptivos, correlacionales, experimentales, etc.), etnográficos, estudios de caso, histórico analíticos, investigaciones evaluativas, colaborativa o participativas [9].

B. Ciencia de Datos

Es la manera como se utiliza el método científico en los datos en un entorno empresarial o investigativo, de esta manera se utiliza matemática, estadística, análisis predictivo, de los datos obtenidos con la finalidad de revelar tendencias, la ciencia de datos es un producto de Big Data desde el inicio hasta el fin y se lo considera como un resultado directo de entornos de datos que se van tornando cada vez más complejos [11].

C. Técnicas de procesamiento y análisis de datos

Los factores decisivos en las técnicas de análisis son: almacenamiento, procesamiento y latencia, por ejemplo, Apache Hadoop recolecta, organiza los datos obtenidos. Para luego utilizar diferentes técnicas de análisis de datos inmediato, como es, la analítica de vídeo, de voz, geoespacial, de lenguaje natural, simulación, modelado predictivo, optimización, extracción de datos y generación de informes, propios de la web 6.0. [10].

III. ADMINISTRACIÓN DE REDES

Según [11], [12], la administración de redes consiste en: aumentar la disponibilidad de la red, mejorar la eficiencia, la seguridad y ayudar al mantenimiento de la red, involucrando objetos que son del nivel más bajo, que componen los entes administrados, que actúan en conjunto con los programas que colectan los datos de administración de las diferentes plataformas y/o elementos de la red.

El agente luego de coleccionar la información, envía a la unidad central de procesamiento, los eventos, con la identidad de los nodos, sus

particularidades y sus diagnósticos en algunos casos. La unidad central de procesamiento es el resultado de aplicaciones centralizadas, que reciben la información enviada por los agentes, almacenan la información, la procesa, y toman acción de las mismas según las políticas establecidas por las unidades de TI [11] [12], como se aprecia en la Figura 1.

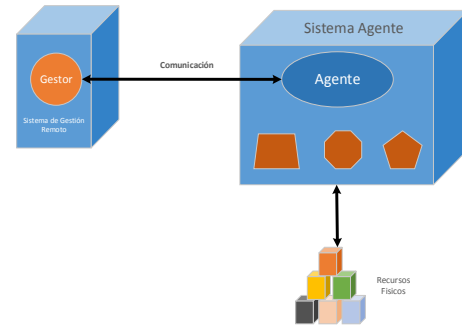


Figura 1. Funcionamiento de Agente en la administración de redes

La Unión Internacional de Telecomunicaciones ITU y Open System Interconnection (Modelo de interconexión de sistemas abiertos) OSI, consideran las actividades que deben ser tomadas en cuenta en el proceso de administración de redes, y concuerdan que son la acumulación de todas las políticas, procedimientos que conforman todas las políticas, procedimientos, que se ejecutan en la organización de las tareas, configuración, control, monitoreo de los componentes que forma parte de la red, con la finalidad de mantener la eficiencia de las comunicaciones entre los nodos, eficiencia que se apreciará en el uso correcto de las aplicaciones [13], [14]. Se ha definido 3 dimensiones para la administración de redes:

- 1) *Dimensión Funcional*- Es asignar las tareas de administración en áreas funcionales;
- 2) *Dimensión Temporal*- Es dividir el proceso de administración en fases repetitivas, considerando las fases de planeación, implementación y operación;
- 3) *Dimensión del escenario*- Es el resto de los escenarios complementarios al de administración de redes.

El modelo se centra en la dimensión funcional Administración de Rendimiento, en su Dimensión Temporal en las fases de planeación.

Así mismo según [13], [14], la administración de rendimiento recolecta y analiza el tráfico, para establecer la conducta en un instante o en un lapso de tiempo, con la finalidad de obtener las mejores decisiones de mejora de *performance*. Esto se realiza en dos etapas: (a) *monitoreo* para recoger la información de la actuación de la red, considerando el porcentaje de transferencia y

recepción de los bits enviados, y luego el (b) *análisis* donde se interpreta la información recogida en el monitoreo, para conocer el comportamiento actual y planificar las acciones necesarias para mantener o mejorar su estado presente.

Principalmente para llevar a cabo la monitorización se utiliza protocolos de red, el SNMP (Simple Network Management Protocol), permite recoger información crítica de la red de una forma estandarizada. Hay varias herramientas que visualizan un mapa de la red sobre los datagramas y otros objetos, como se observa en las Figuras 2, 3.

Las plataformas de monitoreo comprenden un grupo de programas de computador que brindan al administrador servicios para realizar las tareas cotidianas. Algunos ejemplos de suites de gestión actuales son: [15]: HP Open View Network Manager, Sun SunNet Manager, Aprisma Spectrum Site Manager, entre otras.

A. Rendimiento de la Red

Los principales parámetros para medir el rendimiento son: Retraso y Throughput.

- 1) *Retraso (Retardo o Latencia)*- Se produce por la demora en la propagación y transmisión de paquetes en los canales de red [18], [19].
- 2) *Retraso de mensajes*- Se genera debido al tamaño del paquete en conjunto con la rapidez con la cual se propaga por el medio de transmisión. El throughput es la medida de la rapidez de entrega exitosa de una trama, es decir, la capacidad efectiva de transferencia de información del enlace. Por otra parte, el *goodput* da la rapidez de transferencia real, el número de bits de información, que se entregan al destinatario, corresponde al throughput sin los bits de control [18], [19].

Los elementos de claves para determinar el rendimiento de las redes computacionales [13]: Velocidad, Ancho de banda, Seguridad de la red, Confiabilidad, Escalabilidad, Disponibilidad.

- 1) *Velocidad*- Es la tasa de transmisión de datos por segundo en la red.
- 2) *Ancho de banda*- el ancho de banda digital es la capacidad máxima del canal de comunicación. Throughput (ancho de banda efectivo), Goodput (velocidad real

de transmisión de la data (throughput sin las cabeceras)).

- 3) *Seguridad de la red*- las cuatro áreas principales que cubre la seguridad informática [16], [17]: Confidencialidad, Integridad, Disponibilidad, Autenticación.
- 4) *Confiabilidad*- Corresponde a la probabilidad de fallo de un componente de la red, que llevaría a una interrupción del servicio
- 5) *Escalabilidad*- Es la capacidad aumentar o cambiar la red actual ya sea en software o en hardware, pero conservando o mejorando el performance de la red.
- 6) *Disponibilidad*- Es una medida del potencial de la red para permanecer activada y utilizable por los usuarios.

Otro parámetro de desempeño es el CPU de los dispositivos intermedios. Si se duplica la velocidad de CPU, se incrementa el rendimiento del dispositivo, pero una carga extra en el sistema operativo puede generar cuellos de botella en los nodos, requiriendo más CPU, memoria o disco. Incluso actualmente se está tomando en cuenta el aumento de temperatura debido al procesamiento, cuando se ejecutan procesos largos, con el propósito de reducir los costos de operación del sistema.

El tamaño de la ventana (windowing) para reducir la carga extra recomienda enviar unidades de datos grandes para realizar menos envíos y no sobrecargar el canal, como se tendría internamente en un servidor de base de datos, ya que con paquetes pequeños se necesitan más envíos que podrían saturar los buffers [20], [21]. En una red en general la recomendación es: paquetes pequeños con poca cabecera para mejor performance en la transmisión.

Las herramientas más comunes que servirán para diagnosticar correctamente el problema de rendimiento de red se presentan en la Tabla 1, explicando sus características principales y sus plataformas de ejecución [22], [23]:

TABLA 1
PRINCIPALES PAQUETES PARA ADMINISTRACIÓN
DE RED

PAQUETES	PLATAFORMAS	CARACTERÍSTICAS
Netstat	router, linux, windows, swich	Muestra el listado de las conexiones activas
Dropwatch	router, linux, swich	Establece un monitoreo de los paquetes que se envían al canal de comunicación
IP	router, linux, swich	Establece monitoreo de rutas en el tráfico de red
Ethtool	router, linux, windows, swich	Herramienta de configuración de parámetros de red
Snmprsnlate	router, swich	Establece el árbol de la información jerárquica de la base de datos almacenada
Snmplib	router, linux, swich	Comando que permite la administración remota de los hosts.

IV. TÉCNICAS DE ANÁLISIS

A. *Test A / B*: su función es obtener dos versiones de algo y encontrar cual es la mejor. Sirve para generar mejores estrategias de efectividad de landing page, con gran uso en marketing digital [24], [25], [26].

B. *Aprendizaje de reglas de asociación*-, se basa en hallar relaciones entre sus datos, utilizando diferentes heurísticas y/o algoritmos. Como indica [27] es utilizado en grandes bases de datos, como son por ejemplo: Apriori, Eclat y FP-Growth, para luego aplicar técnicas de minería de datos:

- 1) *Algoritmo A priori*- Busca las características comunes de los elementos del conjunto para luego extenderlos a grupos de mayor tamaño con la condición que existan con bastante frecuencia. Es el más conocido y utilizado, destacándose en bases de datos transaccionales porque encuentra de manera rápida y eficaz patrones.
- 2) *Algoritmo de Eclat*- Busca en profundidad los elementos comunes mediante la intersección de los grupos.
- 3) *Algoritmo de FP-crecimiento*- Considerado para patrón frecuente, cuenta cuantas veces se repite los elementos en el conjunto de datos, los almacena, y luego de esto genera

la estructura de árbol FP (patrón frecuente).

- 4) *Clasificación*- Consiste en identificar y categorizar los datos.

C. *Métodos de análisis clúster*. - De acuerdo con [28] una clasificación es la siguiente:

- 1) *Aglomerativa o divisiva* - La aglomerativa inicia con los elementos con una característica común, para luego formar conjuntos, y luego las llamadas particiones. Por otro lado la divisiva o descendente, inicia con todo el conjunto de elementos y luego se los divide progresivamente en conjuntos más pequeños según alguna característica que los diferencie.

- 2) *Monotética o politética*.- Es principalmente la división por separado de los elementos que tienen una característica relevante, generalmente estos casos se ejecutan de manera incremental. [29].

- 3) *Los métodos de optimización* se caracterizan por que admiten la reasignación de los elementos. Esto se da cuando este elemento ya se considera parte del clúster, el siguiente paso de análisis puede o no integrarse a la partición considerándose como más óptimo. Esto es considerado como una mejoría en la partición inicial, normalmente este método asume a priori un número de clúster. Se los denomina de esta manera por que obtiene una partición optimizada de cierta medida definida. Se diferencian otros métodos en como obtienen la partición inicial y de qué manera se optimiza el proceso. Los criterios de optimización suelen ser:

- minimizar la traza de la matriz de variación entre grupos.
- minimizar el determinante matriz de variaciones entre grupos.
- maximizar la traza de la matriz [BW] donde **B** es la matriz de variación entre grupos y **W** es la matriz de variación entre grupos, que presenta similitud de conexión con la discriminación factorial.
- utilizar medidas de estabilidad o de información.

- 4) *Los métodos de densidad*- son basados en la construcción de clúster naturales, iniciando de la mayor a la de menor densidad de variables de las varias zonas del espacio en el que se encuentran los elementos [29].
- 5) *Los métodos clumping*- se utilizan normalmente en los estudios lingüísticos, permitiendo contrastar los grupos, es de ahí que trae su nombre de método de partición[29].

V. TENDENCIA DE LA ADMINISTRACIÓN DE REDES

Tal como lo describe [30], la tecnología en redes ha dado pasos agigantados hacia la transformación digital, por ejemplo informa sobre una empresa que está dando muestras de innovación en este ámbito, en diferentes ámbitos como movilidad, ciberseguridad, cloud computing, análisis de big data, IoT (internet de las cosas), en combinación con estrategias de negocio están impulsando la transformación digital de las grandes y medianas corporaciones y de servicios que se prestan a las personas globalmente.

Expertos en diversos roles dentro de las empresas o corporaciones (CIO (Chief Information Officer), CTO, (Chief Technical Officer)) así como los socios de negocio, proveedores de los servicios, visualizan de una mejor manera una perspectiva más contundente de metas que se cumplieron en el año 2016.

En conjunto con el conocimiento generado con los profesionales, da una mejor visión de las más destacadas fuerzas que modelan y modelarán el cambio en las redes corporativas, durante los próximos años, rediseñando constantemente la administración de: WAN, NFV, IoT, SDN, de la Seguridad, de los servicios basados en localización, VoIP [31], [32].

Como es el caso de una popular empresa de streaming que utilizando los algoritmos de Análisis A/B, logra mejorar la experiencia de televisión por demanda, analizando los datos de sus usuarios y basándose para esto en la ciencia de los datos, en enero de 2016 a nivel global se lanzó a más de 130 nuevos países, convirtiéndose en un portal popular de streaming con un crecimiento planetario, ostenta más de 100 millones de miembros, y aproximadamente la mitad de viven fuera de los Estados Unidos, y almacena datos de sus suscriptores tanto gratuitos como de paga, para entregar contenido segmentado según las preferencias de sus usuarios [33].

Casos de Éxito en la inteligencia artificial son basados en la utilización de Ciencia de Datos, en ámbitos tan indispensables como lo es la protección de datos. [34], salud, retail, el análisis financiero, se basa principalmente en las redes de análisis. La informática juega un papel predominante, y el uso de ciencia de datos es la base de la toma de decisiones en ahora muchos aspectos de la vida cotidiana.

La Tabla 2, relaciona las aplicaciones versus los algoritmos de análisis contemplados en este análisis:

TABLA 2
ANÁLISIS DE HERRAMIENTAS VERSUS UTILIZACIÓN DE ALGORITMOS [35], [36].

	PRTG	HOOT SUITE	Adobe Target	AB TASTY	Optimizely X Web Experimentation	Oracle Enterprise ..	HV OVO
TEST A/B	SI	SI	SI	SI	SI	SI	SI
ANALISIS DE CLUSTER	NO	NO	NO	NO	NO	SI	NO
DISTANCIA MENOR	SI	NO	NO	NO	NO	SI	SI

Entonces se encuentra que la tendencia en el uso de ciencia de datos ocupando todos los algoritmos desarrollados para la explotación parcial o total de la información se está presentando con fuerza en el ámbito de redes sociales, en el mercadeo [37]. Usando en común el algoritmo A/B y el aplicativo con más criterio de análisis se encuentra en Oracle Enterprise Manager [38], contrastado con la publicación de Gartner [39], tal como se aprecia en la Figura 5.

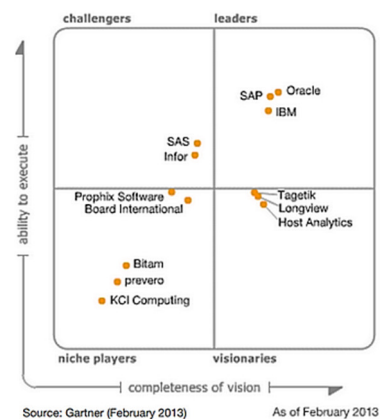


Figura 5. Muestra el cuadrante de Gartner y la posición de Oracle[40]

En la Tabla 3, se relacionan los algoritmos que se utilizan contrastados con los más relevantes dominios:

TABLA 3
ANÁLISIS DE ALGORITMOS USADOS VERSUS LA TENDENCIA DE DOMINIOS APLICADA A LA REDES INFORMÁTICAS [41], [42], [43] [44], [45], [46], [47], [48]

	Detección de Intrusos
Data Mining	Salud Gobierno Poder de Conocimiento Servicios Financieros Detección de Intrusos Retail
Algoritmos de Clasificación	Gobierno Poder de Conocimiento Servicios Financieros Detección de Intrusos Inteligencia Artificial Salud
Machine Learning	Servicios Financieros

ingenierías, salud, gobierno, etc., además se estima que se generan 2.5 trillones de bytes generados por día [37], aumentando la demanda de estos profesionales para este campo, el 32% de los profesiones de análisis de datos son matemáticos y estadistas, el 19% ingenieros informática y el 16% son parte de otras ingenierías [37].

TABLA 4
ANÁLISIS DE TENDENCIAS DE USO DE LA CIENCIA DE DATOS [48]

	Tendencias de Uso
Ciencia de Datos	Negocios
	Redes Sociales
	Económicas
	Deportes
	Ciencias Informáticas

Los centros de estudio especializados en las ciencias de computación, van progresivamente ofertando mallas de estudio para explotar la ciencia de datos. También se evidencia, en el ámbito de las redes de computadoras, que las grandes empresas van ofreciendo análisis autónomos de las redes, poniendo en el mercado aplicaciones que pretenden mejorar la administración de las redes con estos algoritmos [50].

En la información recolectada se ve que hay mayor oferta de soluciones propietarias, evidenciando que falta desarrollar soluciones abiertas o código libre, donde se pueda utilizar sin restricciones los diferentes algoritmos que permitan a los administradores seleccionar el tipo de análisis, creando verdaderos laboratorios de desarrollo de ciencia y tecnología.

El reto principal de la ciencia de datos es brindar una mejor visión de lo que sucedió, lo que sucede y determinar posibles escenarios futuros si la tendencia se mantiene, obteniendo una comprensión gráfica de los patrones y las diferentes relaciones que se presentan entre datos [51], [52].

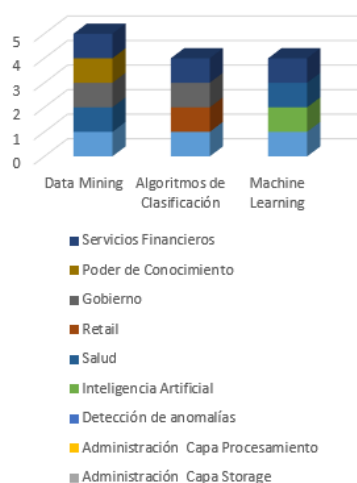


Figura 6. Métodos de Análisis vs Aplicaciones usadas

Los científicos de datos deben conocer de estadística y los datos que se generan a diario potencian perfiles profesionales como son

VI. CONCLUSIONES

La evidencia presentada muestra el interés que existe por obtener la mayor cantidad de información relevante para el análisis de datos que todo día va creciendo ya que la red tiene más presencia en el día a día de las personas, es por este motivo que una parte fundamental para el correcto funcionamiento de toda la red, sea la parte administrativa. Se evidencia que existen aplicaciones que permiten a los administradores obtener la información en sus nodos locales y externos, para cuantificar un sinnúmero de eventos y expresarlos en tableros de control de fácil uso.

Al comparar las aplicaciones evaluadas se observa que el principal algoritmo de análisis de datos es Test A/B, que es una parte fundamental para el análisis más complejo de Clusters. Se ha determinado un cambio en como las empresas van obteniendo más información de sus redes con el objetivo tener el mejor funcionamiento de las mismas. Como parte del análisis también se determinó que el uso de la ciencia de datos no solo se está aplicando para la administración de redes computacionales, sino que también son utilizados en otros ámbitos, por ejemplo, las redes sociales con el objetivo de determinar y conocer gustos, aficiones, interconexiones que pudiese existir con los usuarios de estas, y utilizarlas en tener un mejor impacto en las técnicas de mercadeo que actualmente se utilizan en los medios digitales, y proyectándose incluso a poder llegar a influir en elecciones de un gobierno (Caso Cambridge Analytica) [53].

VII. RECOMENDACIONES

Se recomienda generar aplicativos que utilicen la ciencia de datos, de preferencia con licencia libre, que permita a los administradores de redes, utilizarlos de manera abierta en aspectos que consideren relevantes y sin restricción de criterios en el análisis de las diferentes variables que se puedan presentar.

REFERENCIAS

- [1] C. R. & S. C. 2. Cisco Networking Academy, «Cisco CCNA 2.,» [En línea]. Available: <https://static-course-assets.s3.amazonaws.com/RSE6/es/index.html#1.1.1.1>.
- [2] Babak, «blog.babak.no.» [En línea]. Available: <https://blog.babak.no/en/2007/05/22/why-and-how-to-write-the-state-of-the-art/>.
- [3] I. Xplore, «Machine Learning for Image Reconstruction,» *IEEE*, 2013.
- [4] M. S. F.-e.-b. Saeed Iqbal, «A Machine Learning Based Method for Optimal,» n° ISSN: 1520-9202, 2015.
- [5] reportedigital, «Soluciones Empresariales,» Telefonica, [En línea]. Available: <https://reportedigital.com/cloud/analisis-datos-factor-clave-decisiones/>.
- [6] «Reporte Digital,» [En línea]. Available: <https://reportedigital.com/transformacion-digital/servicios-digitales-grandes-empresas/>.
- [7] techtarget, «techtarget,» [En línea]. Available: <https://www.techtarget.com/es>.
- [8] P. DATA, «Tipos de análisis de datos para la toma de decisiones,» [En línea]. Available: <https://blog.powerdata.es/el-valor-de-la-gestion-de-datos/tipos-de-analisis-de-datos-para-la-toma-de-decisiones>.
- [9] S. Sheerin, «cambridge.org,» [En línea]. Available: <https://www.cambridge.org/core/journals/language-teaching/article/selfaccess/742DE27BD9DD05623CDD6F31594C7365>.
- [10] CISCO, «CISCO WEB6.0,» [En línea]. Available: <https://www.cisco.com/c/en/us/support/customer-collaboration/icm-web-tools/model.html>.
- [11] W. Brad, Big Data Analytics for Dummies, 2015.

- [12] U. Madhow, *Fundamentals of Digital Communications*, Cambridge.
- [13] S. UNTIVEROS, *ADMINISTRACIÓN DE REDES DE COMPUTADORES Conceptos Generales*.
- [14] A. Caicedo, *METODOLOGIAS PARA ADMINISTRAR REDES*, 2004.
- [15] A. Caicedo, «academia,» *ADMINISTRACIÓN DE REDES DE COMPUTADORES*, [En línea]. Available: http://www.academia.edu/11531163/ADMINISTRACION%3%93N_DE_REDES_DE_COMPUTADORES_Conceptos_Generales.
- [16] universidadviu, «Universidad Internacional de Valencia,» 2016. [En línea]. Available: <https://www.universidadviu.es/la-seguridad-informatica-puede-ayudarme/>.
- [17] S. Juliá, «GADAE NET WEB,» 2016. [En línea]. Available: <http://www.gadae.com/blog/5-caracteristicas-red-informatica/>.
- [18] E. Banks, «packetpushers,» 03 2015. [En línea]. Available: <http://packetpushers.net/throughput-vs-goodput/>.
- [19] K. Banger, «.keyboardbanger,» 29 07 2015. [En línea]. Available: <https://www.keyboardbanger.com/throughput-goodput-and-latency/>.
- [20] I. y. Telecomunicaciones, «xuletas,» 2007. [En línea]. Available: <https://www.xuletas.es/ficha/medicion-y-desempeno-de-una-red-lan-1/>.
- [21] C. Armas, «NSRC,» [En línea]. Available: <https://es.slideshare.net/Comdat4/analisis-de-rendimiento-de-red>.
- [22] RedHat, «Red Hat,» [En línea]. Available: https://access.redhat.com/documentation/es-red_hat_enterprise_linux/6/html/performance_tuning_guide/s-network-dont-adjust-defaults.
- [23] linux-party. [En línea]. Available: <https://www.linux-party.com/29-internet/6038-introduccion-a-los-comandos-snmppwalk-snmppget-snmpptranslate>.
- [24] C. Newberry, «HOOTSUITE,» [En línea]. Available: <https://blog.hootsuite.com/es/test-ab/>.
- [25] A. Web, «Digital Guide,» [En línea]. Available: <https://www.1and1.es/digitalguide/online-marketing/analisis-web/testing-tools-programas-para-realizar-tests-ab/>.
- [26] [En línea]. Available: <https://ignsl.es/tecnicas-de-analisis-de-datos/>.
- [27] [En línea]. Available: https://copro.com.ar/Aprendizaje_de_reglas_de_asociacion.html.
- [28] Cuadras, *Introducción al Analisis Cluster. Consideraciones generales..*
- [29] E. S. d. gestión, «ignsl.es,» 2018. [En línea]. Available: <https://ignsl.es/tecnicas-de-analisis-de-datos/>.
- [30] Cisco. [En línea]. Available: https://www.cisco.com/c/es_es/solutions/digital-transformation/index.html.
- [31] CISCO, «CISCO,» [En línea]. Available: https://www.cisco.com/c/dam/global/es_mx/solutions/enterprise-networks/pdf/trends_enterprise_networking_2017_spa.pdf.

- [32] I. T. o. M. Imaging, «IEEE,» 2017. [En línea]. Available: <https://ieeetmi.org/pdfs/TMI-Special-Issue-Machine-Learning-for-Image-Reconstruction.pdf>.
- [33] medium.com. [En línea]. Available: <https://medium.com/netflix-techblog/ab-testing-and-beyond-improving-the-netflix-streaming-experience-with-experimentation-and-data-5b0ae9295bdf>.
- [34] ico.org.uk. [En línea]. Available: <https://ico.org.uk/media/for-organisations/documents/2013559/big-data-ai-ml-and-data-protection.pdf>.
- [35] PRTG, «PRTG,» [En línea]. Available: https://www.es.paessler.com/network-analyzer-diagnostics?utm_source=google&utm_medium=cpc&utm_campaign=ROW_ES_Search-nonBrand_broad_3&utm_adgroup=analysis-de-red&utm_adnum=98878880310&utm_keyword=%2Banalisis%20%2Bde%20%2Bred&utm_device=c&utm_position=1t2&.
- [36] ORACLE, «ORACLE,» [En línea]. Available: <https://go.oracle.com/LP=35729?elqCampaignId=52294&src1=ad:pas:go:dg:bd&src2=wwmk160603p00066c0020&SC=sckw=WWMK160603P00066C0020&mkwid=sOIJbRdB3%7cpcrid%7c105724810964%7cpkw%7can%C3%A1lisis%20de%20datos%7cpmt%7cp%7cpdv%7cc%7csckw=srch:an%C3%A1lisis%20de%2>.
- [37] «xataka,» [En línea]. Available: <https://www.xataka.com/otros/profesion-cientifico-de-datos>.
- [38] Oracle. [En línea]. Available: <https://www.oracle.com/lad/enterprise-manager/index.html>.
- [39] C. Gartner. [En línea]. Available: <https://blog.avanttic.com/2011/11/13/gartner-ratifica-el-liderazgo-de-oracle-en-tecnologias-enterprise-application-server-gobierno-soa-y-enterprise-content-management-en-sus-ultimos-%E2%80%9Ccuadrantes-magicos%E2%80%9D/>.
- [40] gartner. [En línea]. Available: <https://www.gartner.com/doc/reprints?id=1-17GU05Z&ct=110928&st=sb>.
- [41] ADATUM. [En línea]. Available: <https://www.forbes.com/sites/bernardmarr/2015/09/30/big-data-20-mind-boggling-facts-everyone-must-read/#1c87fb3017b1>.
- [42] simplilearn. [En línea]. Available: <https://www.simplilearn.com/data-science-vs-big-data-vs-data-analytics-article>.
- [43] SAS. [En línea]. Available: https://www.sas.com/en_us/insights/analytics/big-data-analytics.html.
- [44] sciencedirect. [En línea]. Available: <https://www.sciencedirect.com/science/article/pii/S0950705116302465>.
- [45] usenix.org. [En línea]. Available: https://www.usenix.org/legacy/publications/library/proceedings/sec98/full_papers/full_papers/lee/lee_html/lee.html.
- [46] V. A. Ayma, «CLASSIFICATION ALGORITHMS FOR BIG DATA ANALYSIS, A MAP REDUCE,» *Dept. of Electrical Engineering, Pontifical Catholic University of Rio de Janeiro, Brazil*.
- [47] B. O. a. M. C. Tarem Ahmed, «Machine Learning Approaches to Network,» *Department of Electrical and Computer Engineering McGill University*.
- [48] SAS, «Bringing the Power of SAS® to Hadoop,» [En línea]. Available:

- https://www.sas.com/en_us/whitepapers/bringing-power-of-sas-to-hadoop-105776.html 1733, June 2016.
doi: 10.1109/TITS.2015.2496783
- URL: <http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=7471487&isnumber=7479588>
- [49] toyo. [En línea]. Available: https://www.toyo.co.jp/files/user/img/contents/ict/pdf/spirent_NA/The_Big_Data_revolution-Is_Your_Network_Ready.pdf.
- [50] «EPN,» [En línea]. Available: <http://www.epn.edu.ec/departamento-de-informatica-y-ciencias-de-la-computacion-dicc/>.
- [51] «marketing4food,» [En línea]. Available: <http://www.marketing4food.com/redes-sociales-usando-ciencia-de-datos/>.
- [52] Cisco, «comstor,» [En línea]. Available: <http://blogmexico.comstor.com/cisco-cree-que-en-el-2020-la-informacion-entre-centros-de-datos-se-triplicara>.
- [53] A. Rodríguez, « Hipertextual,» [En línea]. Available: <https://hipertextual.com/2018/03/que-es-cambridge-analytica>.
- [54] arcgis.com, «arcgis.com,» [En línea]. Available: <https://pro.arcgis.com/es/pro-app/tool-reference/spatial-statistics/optimizedoutlieranalysis.htm>.
- [55] elsevier.com, «elsevier.com,» [En línea]. Available: https://www.elsevier.com/__data/promis_misc/edurevReviewPaperWriting.pdf.
- F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. Di Lorenzo and M. L. Sbodio, "Data-Driven Transit Network Design From Mobile Phone Trajectories," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1724-1733, June 2016.
doi: 10.1109/TITS.2015.2496783
- URL: <http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=7471487&isnumber=7479588>
- G. Di Lorenzo, M. Sbodio, F. Calabrese, M. Berlingerio, F. Pinelli and R. Nair, "AllAboard: Visual Exploration of Cellphone Mobility Data to Optimise Public Transport," in *IEEE Transactions on Visualization and Computer Graphics*, vol. 22, no. 2, pp. 1036-1050, Feb. 1 2016.
- URL: <http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=7117451&isnumber=7368227>
- S. Zhang, D. Yin, Y. Zhang and W. Zhou, "Computing on Base Station Behavior Using Erlang Measurement and Call Detail Record," in *IEEE Transactions on Emerging Topics in Computing*, vol. 3, no. 3, pp. 444-453, Sept. 2015.
doi: 10.1109/TETC.2015.2389614

Referencias Adicionales

F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. Di Lorenzo and M. L. Sbodio, "Data-Driven Transit Network Design From Mobile Phone Trajectories," in *IEEE Transactions on Intelligent Transportation Systems*, vol. 17, no. 6, pp. 1724-

URL: <http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=7008502&isnumber=7225202>

F. Pinelli, R. Nair, F. Calabrese, M. Berlingerio, G. Di Lorenzo and M. L. Sbodio, "Data-Driven Transit Network Design From Mobile Phone

Trajectories," in IEEE Transactions on Intelligent Transportation Systems, vol. 17, no. 6, pp. 1724-1733, June 2016.

doi: 10.1109/TITS.2015.2496783

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=7471487&isnumber=7479588>

C. Dou, Y. Zheng, D. Yue, Z. Zhang and K. Ma, "Hybrid model for renewable energy and loads prediction based on data mining and variational mode decomposition," in IET Generation, Transmission & Distribution, vol. 12, no. 11, pp. 2642-2649, 6 19 2018.

doi: 10.1049/iet-gtd.2017.1476

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8371378&isnumber=8371374>

P. Wang, X. Chen, F. Ye and Z. Sun, "A Smart Automated Signature Extraction Scheme for Mobile Phone Number in Human-Centered Smart Home Systems," in IEEE Access, vol. 6, pp. 30483-30490, 2018.

doi: 10.1109/ACCESS.2018.2841878

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8368227&isnumber=8274985>

A. Khan, I. Ali, A. U. Rahman, M. Imran, Fazal-E-Amin and H. Mahmood, "Co-EEORS: Cooperative Energy Efficient Optimal Relay Selection Protocol for Underwater Wireless Sensor Networks," in IEEE Access, vol. 6, pp. 28777-28789, 2018.

doi: 10.1109/ACCESS.2018.2837108

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8360106&isnumber=8274985>

I. García-Magariño, G. Gray, R. Lacuesta and J. Lloret, "Survivability Strategies for Emerging

Wireless Networks With Data Mining Techniques: a Case Study With NetLogo and RapidMiner," in IEEE Access, vol. 6, pp. 27958-27970, 2018.

doi: 10.1109/ACCESS.2018.2825954

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8345325&isnumber=8274985>

W. Zhao, G. Wang, R. Fan, L. S. Fan and S. Atapattu, "Ambient Backscatter Communication Systems: Capacity and Outage Performance Analysis," in IEEE Access, vol. 6, pp. 22695-22704, 2018.

doi: 10.1109/ACCESS.2018.2828021

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8345348&isnumber=8274985>

W. Zhao, G. Wang, R. Fan, L. S. Fan and S. Atapattu, "Ambient Backscatter Communication Systems: Capacity and Outage Performance Analysis," in IEEE Access, vol. 6, pp. 22695-22704, 2018.

doi: 10.1109/ACCESS.2018.2828021

URL:

<http://bibliotecavirtual.ups.edu.ec:2065/stamp/stamp.jsp?tp=&arnumber=8345348&isnumber=8274985>