ARMADA DEL ECUADOR
ACADEMIA DE GUERRA NAVAL
Guayaquil

TECNOLOGÍAS DE LA INFORMACIÓN EN LA SEGURIDAD Y DEFENSA

Tema:

APLICACIÓN DE LA INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS EN EL ÁMBITO DE SEGURIDAD Y DEFENSA EN EL ECUADOR

Autor

CPCB-IM Milton Mendieta Flores

2018

ÍNDICE DE FIGURAS

ÍNDICE

INTRODUCCIÓN

Los trágicos eventos del 11 de Septiembre del 2011 en los Estados Unidos, EEUU, tuvieron efectos drásticos en muchos aspectos de la sociedad americana que afectaron incluso al resto del mundo. El terrorismo se convirtió en la mayor amenaza a la seguridad nacional debido al gran potencial de daño a la infraestructura, economía y a las personas. En respuesta a estos desafíos, las autoridades federales implementaron en el año 2002 la Estrategia de Seguridad Nacional, NSS, la cual sirvió de referencia para que varios actores estatales, locales y federales estén más vigilantes de todas estas actividades criminales que amenazaban a la seguridad de la nación. La NSS definió seis misiones críticas en donde la dependencia de la Tecnología, principalmente la Inteligencia Artificial, AI, y demás técnicas de ciencia de datos, DS, eran vitales para la consecución de los objetivos de seguridad.

El propósito de este ensayo es analizar la manera cómo los EEUU abordó el cumplimiento de sus objetivos de seguridad nacional planteados en la NSS - 2002 mediante el empleo de tecnologías modernas, como la inteligencia artificial y variadas técnicas de ciencias de datos. Durante el proceso de análisis se propondrán soluciones aplicables en el Ecuador con estas mismas herramientas, orientadas a mejorar el sistema de seguridad nacional, especialmente en el campo marítimo. Para lograr este objetivo se empleará el método inductivo-deductivo, además se utilizará un lenguaje claro y sencillo orientado a lectores sin conocimientos técnicos en las tecnologías mencionadas.

El documento empieza con unas definiciones generales sobre AI y demás técnicas asociadas a la ciencia de datos. Luego se realiza un análisis comparativo sobre la manera que los EEUU aborda las misiones críticas de seguridad establecidas en la NSS - 2002 con AI y ciencia de datos, poniendo especial énfasis en aquellas áreas aplicables a la realidad ecuatoriana en el ámbito de la seguridad y defensa, particularmente en el campo marítimo, utilizando ejemplos reales evidenciados en la literatura científica. En el análisis se incluirán también los desafíos que conlleva el uso de estas tecnologías modernas, siendo el más relevante la restricción legal sobre el uso indiscriminado y no autorizado de información personal. Al final se concluye que el uso de AI y técnicas de DS en el Ecuador, particularmente en el ámbito de la seguridad y defensa, sólo depende de la voluntad de las autoridades ecuatorianas, y una amplia cooperación entre las Fuerzas Armadas, la Academia y el sector privado. Las herramientas disponibles en el mayor de los casos son gratuitas, pero demandan de adecuado conocimiento y personal altamente especializado para logar que nuestras organizaciones tomen decisiones en tiempo real basados en datos.

ANÁLISIS

El terrorismo se convirtió en el enemigo número uno de los EEUU luego de los atentados del 11 de Septiembre, y en respuesta a esta amenaza las autoridades federales elaboraron en el año 2002 la Estrategia de Seguridad Nacional, diseñada para alcanzar tres objetivos de seguridad fundamentales: 1) Prevenir futuros ataques terroristas, 2) Reducir la vulnerabilidad de la nación, y 3) Minimizar los daños y tiempos de respuesta ante la ocurrencia de un ataque (Office of Homeland Security, 2002, p. viii). Las tecnologías de la Información, IT, fueron priorizadas como un elemento indispensable para garantizar la seguridad de la nación (National Research Council, 2002, p. 6). Un estudio realizado en los EEUU identificó varias herramientas de IT que podían contribuir a la obtención de los objetivos de seguridad nacional en cada una de las seis misiones críticas de seguridad (Chen & Wang, 2005), este es el mismo documento que servirá de base para la comparación del caso ecuatoriano.

Antes de continuar con el desarrollo del ensayo, es importante precisar dos asunciones principales. En primer lugar, la NSS de los EEUU utilizada en el presente estudio es la del 2002 y no sus actualizaciones en años posteriores, se considera que este año es una buena referencia de comparación desde el punto de vista tecnológico (AI y DS), en virtud de que el contexto de seguridad que vive el Ecuador en los actuales momentos se encuentra en similares condiciones que los EEUU luego del 911. En segundo lugar, toda referencia relacionada con terrorismo en el caso ecuatoriano, no sólo se limitará a las definiciones clásicas empleadas por los EEUU, sino que también incluirán a todas las actividades derivadas del crimen organizado transnacional: narcotráfico, tráfico de armas, delincuencia organizada, etc, toda vez que la AI y DS también son aplicables en este contexto.

1. DEFINICIONES GENERALES

1.1. CIENCIA DE DATOS

La ciencia de datos es la extracción de conocimiento a partir de grandes volúmenes de información, ya sea estructurada o no estructurada. Una representación gráfica de las áreas que abarca la DS se muestra en el diagrama de Venn de la Figura 1[1], y que actualmente se utiliza en la mayoría de definiciones que se hacen de la materia. La DS trata de una agregación de tres disciplinas esenciales. La primera son las estadísticas y las matemáticas. La segunda es el Conocimiento del Dominio o entorno, por ejemplo: gerencia, publicidad,

[1] El listado de figuras de todo el documento se muestra en el Anexo A.

seguridad y defensa, etc; y, la tercera está relacionada con las habilidades de programación y conocimientos informáticos. Por lo tanto, para que una persona desempeñe el perfil de Científico de Datos, éste debe ser capaz de desempeñar estas tres habilidades.

El Presidente de los EEUU, Barack Obama, le otorgó bastante importancia a los temas relacionados con la DS durante su mandato, tal es así que nombró al primer Chief of Data Scientist, DJ Patil, dentro de la Oficina de Ciencia y Tecnología de la Casa Blanca (Science Friday, 2016). El gobierno creó el repositorio data.gov, donde casi 200.000 sets de datos estaban disponibles para el público de manera gratuita, sirviendo de insumo para que la academia y el sector privado fortalezcan el desarrollo de aplicaciones, productos y/o servicios para los ciudadanos.

1.2. BIG DATA O DATOS MASIVOS

En términos generales, Big Data es un concepto o una herramienta que se utiliza en múltiples disciplinas para describir enormes cantidades de datos (estructurados, no estructurados y semi estructurados) que tomaría demasiado tiempo y sería muy costoso cargarlos a un base de datos relacional para su análisis. El concepto de Big Data aplica para toda aquella información que no puede ser procesada o analizada utilizando procesos o herramientas tradicionales. Sin embargo, Big Data no se refiere a alguna cantidad de datos en específico, ya que es usualmente utilizado cuando se habla en términos de petabytes (10^{15}) y exabytes de datos (10^{18}).

Además del volumen de datos, Big Data opera con una gran variedad de datos, por ejemplo, contenido de redes sociales e internet (Facebook, twitter, contenido web, etc); datos entre máquinas (GPS, lecturas de sensores RFID, lectores Smart, etc); datos generados por los humanos (llamadas celulares, email, historias clínicas digitalizadas), etc. El procesamiento de estos datos demanda de una gran cantidad de computadoras que normalmente están asequibles en la nube, tanto para almacenamiento como para procesamiento de información, el usuario paga sólo lo que consume durante el tiempo que lo consume.

El potencial que ofrece Big Data para las organizaciones, tanto del sector público como el privado, es invaluable. En un reporte elaborado por el Instituto Global McKinsey, se estima que el sector de la salud de los EEUU podría ahorrar más de 300 billones de dólares al año utilizando Big Data de manera creativa y efectiva para mejorar la eficiencia y calidad de los servicios, las dos terceras partes de este valor es equivalente al 8% de gastos del gobierno en salud pública. Similarmente, en las economías desarrolladas de Europa, se estima que la administración gubernamental podría ahorrar más de 100 billones de Euros al año debido a las mejoras en eficiencia operacional al utilizar Big Data (Manyika et al., 2011, p. 2,3). El reporte concluye indicando que los líderes organizacionales a menudo desconocen el valor

que ofrece Big Data, ni tampoco saben cómo explotar este valor. Muchas organizaciones carecen del talento especializado en esta área, y tampoco tienen líderes que sepan tomar decisiones en tiempo real basado en datos.

En el campo militar, la historia es parecida. La denominada IV revolución industrial identifica algunos desafíos para el ámbito de la seguridad y defensa, y propone soluciones tecnológicas como el uso de Big Data: "… optimizar el manejo y la obtención de inteligencia mediante el uso de tecnologías que proporcione una proyección de las capacidades naturales del ser humano, tales como la Big Data" (Gatica, 2017, p. 5).

1.3. MINERÍA DE DATOS

Minería de datos o "data mining" en inglés, se define como el conjunto de técnicas y tecnologías que permiten explorar grandes bases de datos, de manera automática o semiautomática, con el objetivo de encontrar patrones repetitivos, tendencias o reglas que expliquen el comportamiento de los datos en un determinado contexto. En forma general, los datos son la materia prima bruta del proceso de minería de datos. En el momento que el usuario realiza un pre-procesamiento para limpiar el set de datos, escoger el subconjunto de interés, y a éstos atribuirles algún significado especial, los datos pasan a convertirse en información. Cuando los especialistas elaboran o encuentran un modelo, haciendo que la interpretación que surge entre la información y ese modelo represente un valor agregado, entonces nos referimos al conocimiento. Data Mining es una técnica que se encuentra en la intersección de las estadísticas, inteligencia artificial, bases de datos y visualización, según se muestra en la Figura 2.

1.4. INTELIGENCIA ARTIFICIAL

También conocida como Inteligencia Computacional, es una de las ramas de la Informática, con fuertes raíces en otras áreas como la lógica, las ciencias cognitivas, ciencias computacionales, psicología, filosofía, neurociencias, lingüística, análisis operacional, economía, teoría de control, probabilidad y optimización; y tiene como objetivo el estudio y diseño de agentes inteligentes. Un agente inteligente es aquel que percibe el medio donde se desenvuelve, y realiza acciones que maximizan su probabilidad de éxito. Los campos que incluye la Inteligencia Artificial son: redes neuronales o aprendizaje de máquinas, sistemas difusos, computación evolutiva y genética, y sistemas inteligentes.

Un problema complejo como el de pesca ilegal, no regulada y no declarada, IUUF, ha sido abordado por empresas como Schmidt Marine Technology y Vulcan, quienes han desarrollado una plataforma denominada *Skylight* que emplea tecnologías de Big Data y AI

en imágenes satelitales, para determinar con precisión el lugar donde ocurre la pesca ilegal y predecir donde ésta podría ocurrir. Al analizar el movimiento de los buques de pesca, *Skylight* puede identificar qué tipo de pesca está sucediendo y cuándo un buque de pesca parecería estar realizando actividades ilegales (Evans, 2018). El sistema se está utilizando en la actualidad en Palau y en Costa Rica.

1.5. APRENDIZAJE DE MÁQUINAS

También conocido como aprendizaje automático, ML, es una rama de la Inteligencia Artificial (Ver Figura 3), cuyo propósito es lograr que las computadoras actúen sin haber sido explícitamente programadas para el efecto. Los algoritmos utilizados en aprendizaje automático se clasifican en: 1) aprendizaje supervisado, cuando el set de datos está etiquetado, 2) aprendizaje no supervisado, cuando el set de datos no tiene etiquetas, y 3) aprendizaje semi-supervisado que es una mezcla de los dos anteriores. El proceso de aprendizaje de máquinas empieza con el manejo y procesamiento de datos para identificar el subconjunto que será incluido en el modelo de aprendizaje, luego se realiza un proceso de entrenamiento de varios modelos para determinar aquel modelo que mejor aprendió del set de datos, y por último se evalúan modelos con datos no incluidos en el proceso de entrenamiento para escoger el que tenga mejor rendimiento.

1.6. ANÁLISIS DE REDES SOCIALES

El análisis de redes sociales, SNA, es el proceso de investigar estructuras sociales mediante el uso de redes de asociación y teoría de grafos. Las estructuras de redes se representan en términos de nodos (actores individuales, grupo personas, etc), y los enlaces entre los nodos representan las relaciones o interacciones que los conectan (redes de amistad, redes de actividad criminal, redes de transmisión de enfermedades, etc).

1.7. VISIÓN POR COMPUTADORA

Es una disciplina científica que incluye métodos para adquirir, procesar, analizar y comprender las imágenes del mundo real con el fin de producir información numérica o simbólica para que puedan ser tratados por un computador. Tal y como los humanos usamos nuestros ojos y cerebros para comprender el mundo que nos rodea, la visión por computadora trata de producir el mismo efecto para que las computadoras puedan percibir y comprender una imagen o secuencia de imágenes y actuar según convenga en una determinada situación. Algunas técnicas empleadas en el campo de la visión por computadora son: reconocimiento

facial, detección de objetos, clasificación de imágenes, etc. La detección de objetos en imágenes o en video consiste en identificar los objetos de interés en la escena y colocar un rectángulo a su alrededor, según se aprecia en la Figura 4. El reconocimiento facial se basa en algoritmos que identifican el rostro de una persona en una imagen digital, normalmente al comparar las características más sobresalientes de la imagen, con imágenes disponibles en una base de datos, según se observa en la Figura 5.

2. MISIONES CRÍTICAS DE SEGURIDAD

La NSS definió 6 misiones críticas de seguridad; sin embargo, el ensayo está orientado únicamente a las cuatro misiones que están más relacionadas con las actividades realizadas por la Armada del Ecuador.

2.1. INTELIGENCIA Y VIGILANCIA

Aunque los terroristas dependen de la sorpresa para realizar sus ataques, sus actividades no son aleatorias, ni tampoco son difíciles de traquear. Los terroristas deben planificar antes de ejecutar el ataque, para lo cual seleccionan el objetivo, reclutan y entrenan a las personas encargadas de materializar el ataque, consiguen apoyo financiero y viajan al país donde está ubicado su objetivo. Para evitar ser detectado por las autoridades, los terroristas tratan de ocultar sus identidades verdaderas, y también disfrazan sus actividades relacionadas con el ataque de la misma manera que otros criminales lo hacen.

Una técnica para detectar individuos y organizaciones dedicadas a actividades ilegales son las redes de actividad criminal, CAN, estas redes se usan para analizar información de múltiples fuentes como las agencias de seguridad, tránsito, etc, no sólo de una jurisdicción sino de múltiples jurisdicciones. Una CAN es una red de gente interconectada (criminales conocidos), vehículos y localidades basadas en registros policiales y judiciales (Kaza, Xu, Marshall, & Chen, 2009), un ejemplo se muestra en la Figura 6. Utilizando teoría de grafos, se puede analizar los “cliques”, que representan círculos de amigos y conocidos que forman redes sociales. Utilizando una métrica conocida como “clustering coefficient”, se puede medir la tendencia de individuos para cooperar y agruparse en dichos cliques para el cometimiento de delitos. Otra métrica utilizada en grafos es el “degree distribution”, que mide la cantidad de enlaces que llegan a un solo nodo, un valor alto en una CAN implica el liderazgo que puede tener un nodo específico en una red determinada.

Este mismo enfoque puede ser aplicado en la institución, en cooperación con las demás agencias del Estado, para elaborar redes de actividades criminales, incluso se puede utilizar la información de personal que en el pasado se ha involucrado en actividades

delictivas, para determinar posibles cooperaciones con personal naval en servicio activo, e identificar al resto de miembros de bandas delictivas. Una fuente muy importante de información para mejorar la robustez de las CAN son los detalles de llamadas telefónicas, aunque esta información se consigue sólo con autorización judicial.

Las tecnologías de visión por computadora y AI son utilizadas actualmente en tareas de vigilancia y exploración aeromarítima, EAM, tal es el caso del primer radar óptico ViDAR (Visual Detection and Ranging) desarrollado por la empresa británica SENTIENT. El ViDAR consiste de una o varias cámaras COTS[2] de 9 megapixeles, funciona análogamente como un radar pero en el dominio de la visión (SENTIENT, s. f.), según se observa en la Figura 7. Los objetos detectados por las cámaras se envían al operador para su análisis posterior mediante otros sensores a bordo de la aeronave. Esta aplicación, de bajo costo en relación a la adquisición de un radar y/o equipo electroóptico, EO, puede ser desarrollada en el Ecuador para implementarse en la Aviación Naval, ya sea como un mecanismo de detección secundario en aquellas aeronaves donde ya se cuente con equipos EO, esto es aviones de EAM tipo CASA y Super King Air; o como un mecanismo de detección principal en aeronaves ligeras tipo CESSNA y PILLÁN que realizan EAM costera, incluso en los helicópteros livianos BELL-206 o TH-57.

Las aeronaves y helicópteros ligeros tienen la restricción de instalación de equipos pesados como un radar y/o EO, ya que no disponen de puntos de fijación fuertes en el fuselaje, además que el peso adicional reduce aún más su limitada autonomía; por lo que la alternativa de un sistema como el ViDAR es viable en estos casos. Un concepto similar se puede implementar en futuros UAV´s tácticos empleados por unidades de Infantería de Marina para detectar objetos pequeños en movimiento, según se muestra en la Figura 8. En ambos casos, los videos e imágenes pueden ser post procesados con algoritmos de compresión para ser transmitidos desde las aeronaves hacia estaciones de control en tierra en tiempo real mediante los enlaces de datos disponibles.

2.2. SEGURIDAD EN FRONTERAS Y TRANSPORTE

Los terroristas entran al país objetivo por aire, mar y tierra, donde las autoridades aduaneras y de inmigración colectan información diariamente, tales como: identificación de viaje, imágenes, huellas digitales, vehículos utilizados, entre otros. Al intercambiar y analizar información de múltiples fuentes se pueden crear “fronteras inteligentes”, estas fronteras dependen de tecnologías tales como colaboración, comunicación, biometría, y reconocimiento de imágenes y voz (Chen & Wang, 2005, p. 13).

[2] COTS.- Commercial-of-the-shelf, está relacionado con artículos que no son de estándar militar

Una aplicación relacionada con lo especificado en el párrafo anterior es el sistema de predicción de crimen denominado “PredPol”. Un grupo de estudiantes en la Universidad de California (UCLA) desarrolló un software de predicción de cometimiento de ilícitos utilizando AI y Big Data”, y fue probado en el Departamento de Policía de Los Ángeles, cuyos resultados fueron muy satisfactorios en lo relacionado a la reducción del crimen. El software envía a las patrullas un área de 500 pies cuadrados donde se presume se cometerán ilícitos, según se muestra en la Figura 9. Este software que nació como un proyecto universitario, ahora se comercializa a nivel global (PredPol, s. f.).

Una aplicación similar en el caso ecuatoriano se podría desarrollar en la Armada del Ecuador. La DIRNEA cuenta con el Sistema de Gestión Marítimo Portuario (SIGMAP). Este sistema tiene automatizado ciertos procesos de la entidad: 1) Procesos de Autoridad Marítima: personal mercante, registro de naves, inspección de naves, documentos vigentes, etc; 2) Permiso de tráfico: sin este permiso no se puede otorgar el zarpe; 3) Sistema de Monitoreo Satelital: georreferenciación de rutas y posición de embarcaciones mayores a 20 TRB; 4) Interconexión con los sistemas informáticos de Petroecuador para la venta de combustible en base al consumo calculado por la distancia y velocidad calculada en el sistema de monitoreo satelital; 5) Georreferenciación de delitos. Con toda esta información, y agregando datos complementarios de otras fuentes, se pueden diseñar modelos de predicción del cometimiento de ilícitos utilizando técnicas de data mining, Big Data, machine learning y AI, similar a PredPol, permitiendo que la Armada del Ecuador optimice el patrullaje de sus unidades navales y guardacostas en aquellas áreas identificadas en tiempo real como potenciales zonas de riesgos. No sólo se trata de georreferenciar el crimen sino prevenirlo.

El sistema PredPol – SIGMAP propuesto puede incluso ser complementado con información en tiempo real proveniente de cámaras de video provistas al personal embarcado en las lanchas guardacostas. Las imágenes o videos capturados permitirán extraer el nombre de la embarcación de manera automática para determinar su status legal, todo esto mientras la lancha está en movimiento; además de buscar personas con antecedentes delictivos mediante reconocimiento facial de los ocupantes de las embarcaciones registradas.

2.3. CONTRA TERRORISMO DOMÉSTICO

Los expertos consideran al terrorismo como un tipo de organización criminal similar a las pandillas o narcotraficantes, en donde múltiples criminales cooperan en el cometimiento de delitos. La IT puede encontrar relaciones cooperativas y patrones de interacción entre criminales, utilizando técnicas de SNA similares a la descrita en el numeral 2.1. Parte del contra terrorismo doméstico, o en el caso ecuatoriano la lucha contra el crimen organizado en el país, es contar con personal militar y policial comprometido con la institución, y que no esté

involucrado en actividades ilícitas. En la Armada del Ecuador se han realizado pequeños esfuerzos utilizando técnicas de minería de datos para elaborar un modelo descriptivo que permita identificar al personal naval propenso a cometer faltas disciplinarias (Mendieta, 2018). Se espera en el corto plazo utilizar los hallazgos en dicho estudio preliminar para diseñar un modelo de predicción de cometimiento de faltas disciplinarias en base a los datos existentes en la Dirección de Talento Humano, DIGTAH, de la Armada del Ecuador. Si bien es cierto que las faltas disciplinarias son de carácter administrativo y no penal, este dominio del conocimiento puede ser extrapolado hacia la predicción e identificación de personal con alta probabilidad de incurrir en actividades delictivas, además de sus redes de cooperación, usando técnicas de AI y minería de datos.

2.4. PROTECCIÓN DE INFRAESTRUCTURA FÍSICA Y ACTIVOS CLAVES

La infraestructura física vital de un país se convierte en objetivos potenciales para los ataques terroristas. La infraestructura virtual como la internet también es vulnerable a este tipo de amenazas. Para monitorear estos activos claves, no sólo se requiere de sensores y mecanismos de detección, sino que también se requiere de tecnologías avanzadas de información que puedan modelar comportamientos normales y puedan al mismo tiempo detectar y diferenciar los comportamientos anormales. Estos modelos se realizan utilizando técnicas de ML, Big Data y AI. La ciber defensa se encarga de proteger la infraestructura de redes y de internet. Todos los datos que se recolectan de manera permanente durante el monitoreo de las redes e infraestructura de internet, deben ser modelados para detectar amenazas desconocidas (zero-day attacks) para lo cual no existen mecanismos de detección.

Una aplicación en el Ecuador, particularmente en la Armada, sería utilizar técnicas de visión por computadora, procesamiento de imágenes y ML para detectar objetos (personas, vehículos, rostros, etc) en los sistemas de monitoreo y vigilancia de los repartos navales, esto le permite al operador de video cámaras canalizar su atención hacia aquellas imágenes de interés que de otra manera pasarían desapercibidas.

Otra aplicación práctica es utilizar una técnica similar a la descrita en el párrafo anterior para proteger los pañoles de artillería de robos de armamento. Con técnicas de detección de rostro se identifica desde las cámaras de video del pañol al personal que retira armamento. Usando algoritmos de clasificación con ML, con las mismas imágenes de las cámaras de seguridad se identifica el armamento que requiere el personal, y con una cámara de mayor resolución se lee la serie del fusil usando técnicas de segmentación de caracteres, según se muestra en la secuencia descrita en la Figura 10. La información del personal, armamento, y número de serie se registra en una base de datos para mantener el control de inventarios en tiempo real. Lo único que se requiere para implementar esta propuesta es tomar la foto de

todo el personal naval para entrenar los algoritmos de detección de rostros, y varias fotos de cada arma disponible en el pañol para entrenar los algoritmos de clasificación y segmentación de caracteres.

El uso de técnicas de AI y DS provee gran valor a las organizaciones, incluso a las agencias e instituciones dedicadas a la seguridad y defensa. El Sr. VALM (SP) Marco Salinas en su ensayo "Guerras del Futuro: Causas y Estrategias" mencionaba que en el año 2030 se iba a configurar en el mundo la tormenta perfecta, y que para esa fecha la población mundial estaría navegando en un mar de incertidumbres, en primera clase estarán los países que se encuentran en la ola del conocimiento (Salinas, 2018, p. 8), justamente aquellos que han desarrollado tecnologías basadas en la AI y Ciencia de Datos.

Todos los ejemplos y/o aplicaciones propuestas en el numeral 2 del presente documento tienen potencial para ser desarrollados en el Ecuador, mediante un esfuerzo coordinado entre las FF.AA. – Armada del Ecuador, academia y el sector privado. Las FF.AA. porque conocen el proceso y en la mayoría de los casos disponen de los datos para elaborar modelos de AI; la academia porque dispone del talento altamente especializado para abordar estos problemas complejos; y por último el sector privado donde se madurarán estas tecnologías con potencial para ser aplicadas al ámbito comercial, contribuyendo al cambio de la matriz productiva del país. Los datos están disponibles en INOCAR, DIRNEA, COGUAR, Aviación Naval, DIGTAH, entre otros repartos navales; simplemente no han sido procesados para extraer conocimiento valioso para la institución. Las herramientas para el procesamiento de estos modelos están disponibles de manera gratuita en internet, un detalle se muestra en el Anexo B. La Armada del Ecuador en su nuevo plan de gestión "BICENTENARIO", ha visualizado la necesidad de implementar en el Comando de Operaciones Navales un Centro de Análisis de Inteligencia Naval y un Centro Naval de Ciber Operaciones y Operaciones de Información (ESMAAR, 2018). Estos nuevos centros obligatoriamente deberán fundamentar su operación en el empleo de técnicas de AI, DS y Big Data, caso contrario su aporte a la institución tendrá escaso valor; toda vez que la Inteligencia, ciber operaciones y operaciones de información se sustentan en el uso masivo de datos.

3. DESAFÍOS DE LA INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS EN EL DOMINIO DE LA SEGURIDAD Y DEFENSA

Toda acción encaminada a la lucha contra el crimen organizado trae consigo problemas únicos de IT, así como también sus propios desafíos. Las organizaciones criminales tienen un comportamiento distribuido, es decir, están dispersas geográfica y temporalmente. Como resultado, las investigaciones deben cubrir a múltiples involucrados y actividades criminales en diferentes lugares y en tiempos diferentes. Esta es una situación

muy compleja debido a los escasos recursos disponibles en las agencias de inteligencia y seguridad. A medida que avanzan las tecnologías en computación e internet, los criminales explotan el ciberespacio para cometer varios tipos de ciber crímenes disfrazados como transacciones y comunicaciones online ordinarias.

En segundo lugar tenemos que el dominio de la inteligencia, seguridad y defensa no dispone de una fuente de datos bien definida, esto significa que los investigadores deben obtener no sólo información oficial (partes policiales, registros telefónicos, estados de cuenta bancarios, registros de inmigración, etc), sino también información de fuente abierta (noticias, artículos científicos, libros, páginas web, etc). Los formatos de los datos varían desde una base de datos estructurada hasta datos no estructurados como el texto, imágenes, audio y archivos de video. Las asociaciones criminales pueden ser extraídas sólo de los datos no estructurados, los cuales son muy difíciles de obtener (Chen & Wang, 2005, p. 14).

Otro desafío importante está relacionado con las técnicas disponibles para el análisis de delitos e inteligencia. La IT ha desarrollado varias herramientas y metodologías, la mayoría son gratuitas, relacionadas con la integración de datos, análisis de datos, minería de texto, procesamiento de imágenes y video, etc; sin embargo, el empleo eficiente de estas herramientas en el dominio de la inteligencia, seguridad y defensa continúa siendo una pregunta sin responder. Transformar los datos crudos, obtenidos del sistema de seguridad nacional, en información de inteligencia útil para la toma de decisiones, requiere invertir sustancialmente en investigación en varias sub disciplinas de la AI: minería de datos, texto y web; procesamiento natural de lenguaje; planeamiento; resolución de conflictos, análisis de enlaces, y algoritmos de búsqueda, etc (Chen & Wang, 2005, p. 14).

En virtud de todos los desafíos tecnológicos descritos anteriormente, la comunidad científica internacional organiza anualmente una conferencia sobre Informática de Inteligencia y Seguridad, ISI, cuyo objetivo es desarrollar tecnologías avanzadas de información, sistemas, algoritmos, y bases de datos orientadas al empleo de aplicaciones de seguridad nacional, mediante un enfoque tecnológico y organizacional. Es importante mencionar que en el Ecuador se realizó el I congreso internacional MICRADS (Multidisciplinary International Conference of Research Applied to Defense and Security) en el mes de Abril 2018 en la Escuela Superior Naval “Comandante Rafael Morán Valverde”, orientado al ámbito de la seguridad y defensa (MICRADS, s. f.). Este congreso es organizado por la ESPE cada dos años, y deja muy en claro que en el país las autoridades se están dando cuenta del potencial uso de las tecnologías para la resolución de problemas militares y de seguridad.

En último lugar está el desafío más importante de todos, relacionado con el aspecto legal del uso de la información. Aunque el análisis legal está fuera del alcance del presente estudio, es importante mencionar que en el Ecuador existen varias leyes y normas que regulan todo los aspectos relacionados con la colección de datos, confidencialidad y reporte de la

información, los cuales pueden tener un impacto negativo directo en el desarrollo de aplicaciones de AI orientadas al ámbito de la inteligencia, seguridad y defensa. El Artículo 20 de la Ley de Seguridad Pública y del Estado establece los casos cuando se requiere autorización judicial para obtención de información (Asamblea Nacional, 2009). No es ético ni legal “pescar al azar” en búsqueda de potenciales criminales utilizando la información disponible en varias fuentes: inteligencia, datos de los ciudadanos, repositorios de datos de crímenes, etc. Los investigadores deben siempre utilizar un enfoque basado en hipótesis o evidencia, esto significa que deben existir causas probables o razonables y evidencia, previo al análisis de individuos o un set de datos en particular (Chen & Wang, 2005, p. 15). El Congreso de los EEUU cerró el programa MATRIX del Departamento de Defensa, por considerar que existía potencial para el mal uso de la información de los ciudadanos, afectando sus derechos civiles.

CONCLUSIONES

Con los argumentos fundamentados en los párrafos anteriores, y tomando en consideración las soluciones tecnológicas propuestas en el ámbito de la seguridad y defensa, a continuación se establecen las siguientes conclusiones:

a. El empleo de tecnología moderna como la Inteligencia Artificial y Ciencia de Datos en el ámbito de la seguridad y defensa en el Ecuador, permitirá resolver problemas complejos desde un enfoque de datos para tomar decisiones en tiempo real.

b. Los datos existentes en las instituciones del sistema de seguridad y defensa del Ecuador, específicamente la Armada del Ecuador, sumado al talento especializado disponible en la Academia y el sector privado, y considerando que las herramientas de desarrollo están disponibles de manera gratuita en el internet, permitirán el desarrollo de las soluciones tecnológicas propuestas en el presente documento, colocando al sector de seguridad y defensa a la vanguardia de la lucha contra el crimen organizado.

c. La concienciación de los líderes navales sobre el valor que tiene para la organización tomar decisiones en base a datos, facilitará a la institución transitar por el sendero del conocimiento, mejorando así las capacidades para enfrentar los desafíos y amenazas del futuro.

LISTADO DE SIGLAS Y ACRÓNIMOS

AI.- Artificial Intelligence – Inteligencia Artificial

CAN.- Criminal Activity Network – Redes de Actividad Criminal

DS.- Data Science – Ciencia de Datos

IUUF.- Illegal, unregulated and unreported fishing – Pesca Ilegal, No Regulada y No Declarada

ISI.- Intelligence and Security Informatics – Informática de Seguridad e Inteligencia

MATRIX.- Multistate Antiterrorism Information Exchange – Intercambio Multiestado de Información Antiterrorista.

MICRADS.- Multidisciplinary International Conference of Research Applied to Defense and Security – Conferencia Multidisciplinaria Internacional de Investigación Aplicada a la Defensa y la Seguridad

ML.- Machine Learning – Aprendizaje de Máquinas

NSS.- National Security Strategy – Estrategia de Seguridad Nacional

SNA.- Social Network Analysis – Análisis de Redes Sociales

BIBLIOGRAFÍA

Asamblea Nacional. (2009, septiembre). Ley de Seguridad Pública y del Estado. Recuperado 23 de abril de 2018, a partir de https://www.google.com/search?q=ley+de+seguridad+publica+y+del+estado&ie=utf-8&oe=utf-8&client=firefox-b-ab

Chen, H., & Wang, F.-Y. (2005). Artificial intelligence for homeland security. *IEEE Intelligent Systems*, *20*(5), 12–16.

ESMAAR. (2018, marzo 2). Instructivo Plan Gestión Institucional «BICENTENARIO» Directrices.

Evans, I. (2018, febrero 14). Deeply Talks: Fighting Illegal Fishing With Big Data, Robots and A.I. Recuperado 11 de abril de 2018, a partir de https://www.newsdeeply.com/oceans/articles/2018/02/14/deeply-talks-fighting-illegal-fishing-with-big-data-robots-and-a-i

Gatica, J. (2017). La 4ta. Revolución Industrial y su Impacto en la Seguridad y Defensa. *Centro de Investigaciones y Estudios Estratégicos*.

Kaza, S., Xu, J., Marshall, B., & Chen, H. (2009). Topological analysis of criminal activity networks: Enhancing transportation security. *IEEE Transactions on Intelligent Transportation Systems*, *10*(1), 83–91.

Manyika, J., Chui, M., Brown, B., Bughin, J., Dobbs, R., Roxburgh, C., & Byers, A. H. (2011, mayo). Big data: The next frontier for innovation, competition, and productivity | McKinsey & Company. Recuperado 18 de febrero de 2018, a partir de https://www.mckinsey.com/business-functions/digital-mckinsey/our-insights/big-data-the-next-frontier-for-innovation

Mendieta, M. (2018). Generating a Descriptive Model to Identify Military Personnel Incurring in Disciplinary Actions: A Case Study in the Ecuadorean Navy (Pending publication). *MICRADS 2018*.

MICRADS. (s. f.). Multidisciplinary International Conference of Research Apllied to Defense and Security. Recuperado 18 de febrero de 2018, a partir de http://www.micrads.org/index.php?lang=en

National Research Council. (2002). *Making the nation safer: The role of science and technology in countering terrorism*. National Academies Press.

Office of Homeland Security. (2002, noviembre 1). National Strategy For Homeland Security. Recuperado 17 de febrero de 2018, a partir de https://www.dhs.gov/national-strategy-homeland-security-october-2007

PredPol. (s. f.). Predict Crime | Predictive Policing Software. Recuperado 18 de febrero de 2018, a partir de http://www.predpol.com/

Salinas, M. (2018). Guerras del Futuro: Causas y Estrategias (publicación pendiente). Presentado en MICRADS 2018, Salinas, Ecuador.

Science Friday. (2016, abril 19). 10 Questions for the Nation's First Chief Data Scientist. Recuperado 18 de febrero de 2018, a partir de https://www.sciencefriday.com/articles/10-questions-for-the-nations-first-chief-data-scientist/

SENTIENT. (s. f.). ViDAR. Recuperado 11 de abril de 2018, a partir de http://www.sentientvision.com/products/vidar/

ANEXO A

LISTA DE FIGURAS

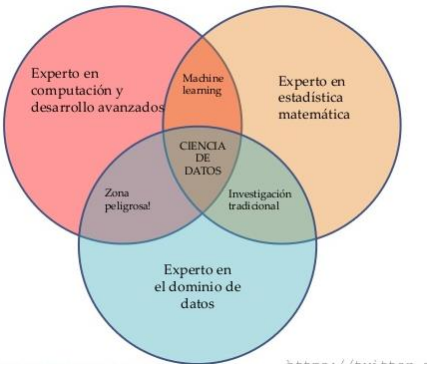

Figura 1. Áreas que abarca la Ciencia de Datos
Fuente: http://fc5scrim.blogspot.com/2015/08/los-cientifico-de-datos.html

Figura 2. Intersección de la Minería de Datos con otras áreas del conocimiento
Fuente: http://pruebita.url.ph/index.php/bolsa-de-trabajo/12-blog/34-la-evolucion-hasta-llegar-a-la-mineria-de-datos

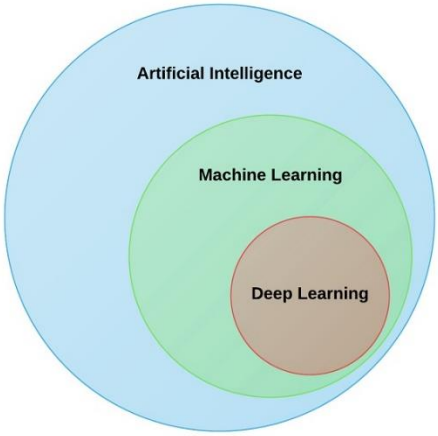

Figura 3. Relación entre la Inteligencia Artificial y el Aprendizaje de Máquinas
Fuente: https://aitrends.com/machine-learning/artificial-intelligence-vs-machine-learning/

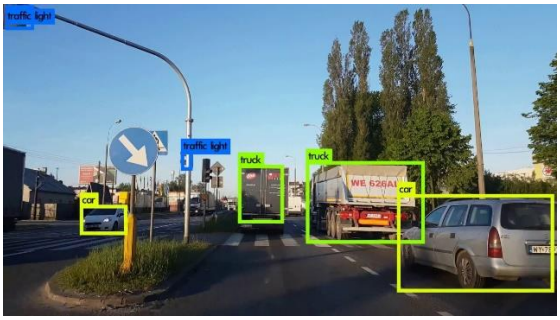

Figura 4. Detección de Vehículos en video

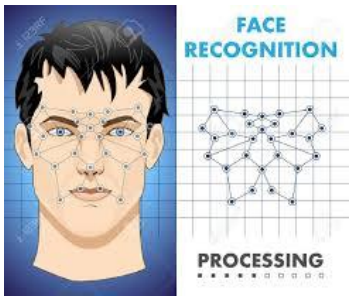

Figura 5. Reconocimiento Facial
Fuente: https://www.123rf.com/photo_47856665_stock-vector-face-recognition-biometric-security-system.html

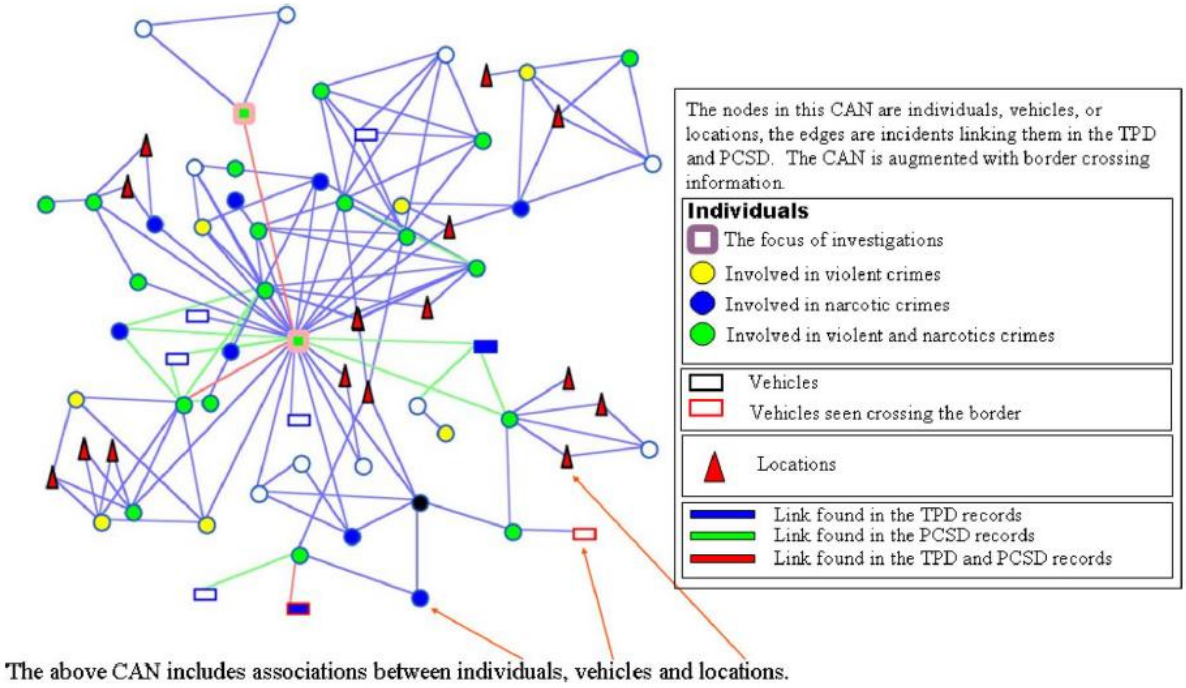

Figura 6. Ejemplo de una red de actividad criminal (CAN) relacionada con narcóticos
Fuente: Topological Analysis of Criminal Activity Networks, IEEE Transactions on Intelligent Transportation Systems, Vol. 10, No.1, March 2009

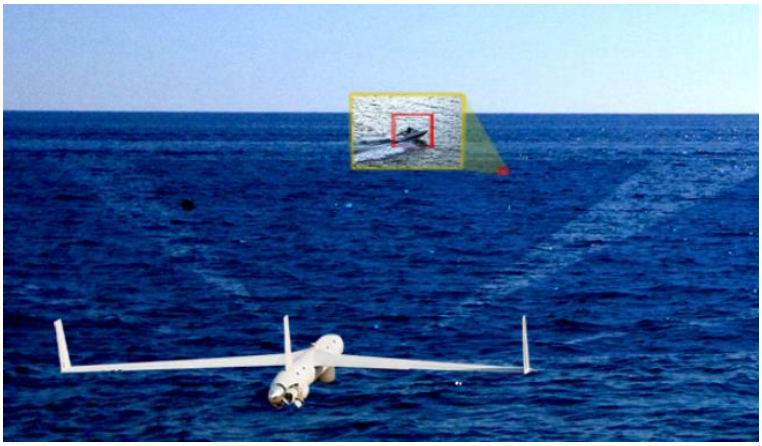

Figura 7. Detección de contactos por medio de cámaras de video y algoritmos de Inteligencia Artificial
Fuente: http://www.sentientvision.com/products/vidar/

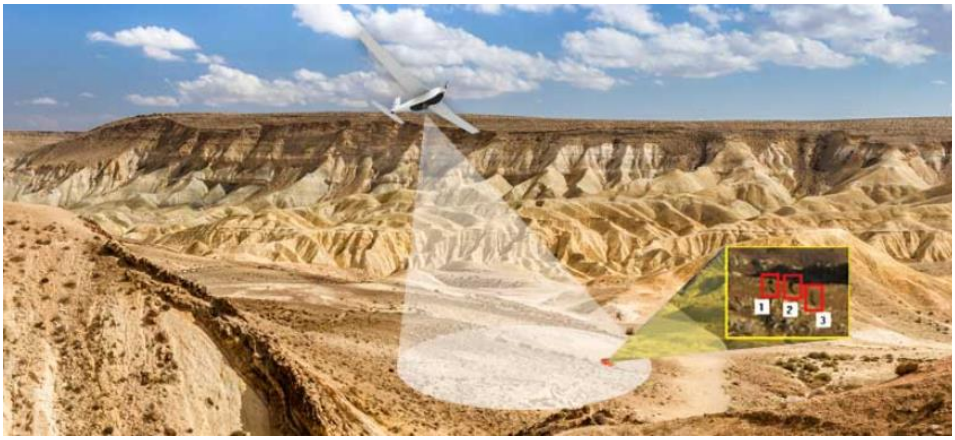

Figura 8. Detección de objetos pequeños en movimiento en tierra
Fuente: http://www.sentientvision.com/products/kestrel-land-mti/

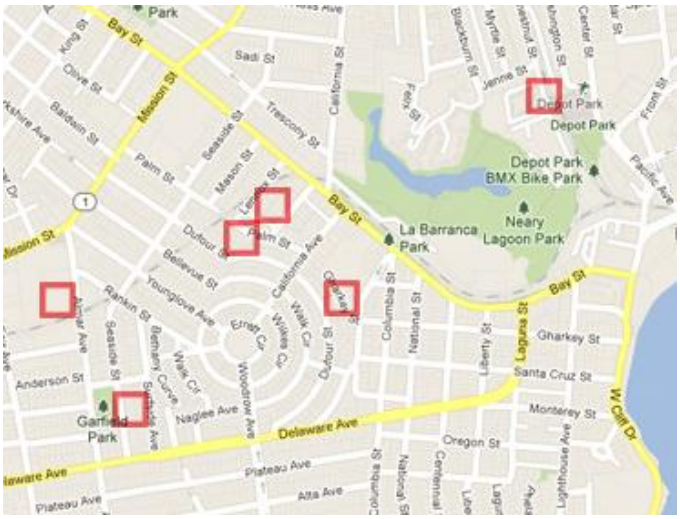

Figura 9. Sistema de Predicción de Crímenes “PredPol”
Fuente: www.predpol.com

Figura 10. Pañol de Artillería del CUINMA*. Izquierda:* vista frontal del pañol. *Centro:* Reconocimiento facial con cámara. *Derecha:* reconocimiento del arma y número de serie

ANEXO B

HERRAMIENTAS OPEN-SOURCE GRATUITAS PARA CIENCIA DE DATOS, BIG DATA, VISIÓN POR COMPUTADORA E INTELIGENCIA ARTIFICIAL

1. **Python.-** es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Es un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional que se emplea predominantemente en big data. Es un lenguaje interpretado, usa tipado dinámico y es multiplataforma. Puede ser descargado de su página oficial https://www.python.org/downloads/ para varios sistemas operativos: Windows, Linux/Unix, Mac OS X y otros.
2. **Numpy.-** es una extensión de Python que le agrega mayor soporte para vectores y matrices, constituyéndose en una biblioteca de funciones matemáticas de alto nivel para operar con dichos vectores o matrices. Puede ser descargado de su sitio oficial http://www.numpy.org/, es multiplataforma por lo que puede ser instalado en varios sistemas operativos: Windows, Linux/Unix, Mac OS, entre otros.
3. **Pandas.-** es una librería de Python escrita como extensión de NumPy para manipulación y análisis de datos. En particular, ofrece estructuras de datos y operaciones para manipular tablas numéricas y series temporales. Puede ser descargado desde su página oficial https://pandas.pydata.org/pandas-docs/stable/install.html, es multiplataforma al igual que NumPy.
4. **Scikit-Learn.-** es una librería de Aprendizaje de Máquinas utilizada con Python. Está diseñada para interoperar con las libreríass numéricas y científicas NumPy y SciPy. Está disponible para sistemas operativos Windows, Linux/Unix, Mac OS, y puede ser descargada de su sitio oficial http://scikit-learn.org/stable/install.html.
5. **TensorFlow.-** es una librería de código abierto utilizada en el campo de Aprendizaje de Máquinas o Aprendizaje Automático. Fue desarrollada por Google para satisfacer sus necesidades de sistemas capaces de construir y entrenar redes neuronales para detectar y descifrar patrones y correlaciones, análogos al aprendizaje y razonamiento usados por los humanos. TensorFlow fue originalmente desarrollado por el equipo de Google Brain para uso interno en Google antes de ser publicado bajo la licencia de código abierto Apache 2.0 el 9 de noviembre de 2015. Puede ser descargado de su sitio oficial https://www.tensorflow.org/install/, tanto para Windows, Linux/Unix, o Mac OS.
6. **Keras.-** es una librería de código abierto escrita en Python, capaz de correr sobre varios motores de procesamiento tales como: TensorFlow, Theano, MXNet, entre otros. Fue diseñada para permitir experimentos rápidos con redes neuronales profundas, su enfoque

es ser amigable con el usuario, modular y extensible. Puede ser descargada de su sitio oficial https://keras.io/.

7. **OpenCV.-** Es una librería de código abierto de visión artificial originalmente desarrollada por Intel. Open CV es multiplataforma, existiendo versiones para GNU/Linux, Mac OS X y Windows. Contiene más de 500 funciones que abarcan una gran gama de áreas en el proceso de visión, como reconocimiento de objetos (reconocimiento facial), calibración de cámaras, visión estérea y visión robótica. Puede ser descargada de su sitio oficial http://opencv.org/.
8. **NetworkX.-** es una librería escrita en Python utilizada para análisis de redes y grafos, que se usan extensivamente en análisis de redes sociales. Puede ser descargada del siguiente link https://pypi.org/project/networkx/2.1/.
9. **Apache Hadoop.-** es un framework de Big Data que soporta aplicaciones distribuidas bajo una licencia libre. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop está compuesto por su paradigma de programación MapReduce y el sistema distribuido de archivos llamado Google File System (GFS). Sobre estos dos motores se ha montado todo un ecosistema de variadas aplicaciones para solucionar problemas específicos de Big Data, tales como: Apache Pig que es un lenguaje de scripting para manejar información estructurada, Apache Hive para realizar consultas tipo SQL, Mahout para realizar Aprendizaje de Máquinas, entre otros. Puede ser descargado de su página oficial http://hadoop.apache.org/releases.html, es multiplataforma.
10. **Apache Spark.-** Se puede considerar como un sistema de computación en cluster de propósito general y orientado a la velocidad. Proporciona APIs en Java, Scala, Python y R. También proporciona un motor optimizado que soporta la ejecución de grafos. Al igual que Hadoop, soporta todo un ecosistema o conjunto extenso y rico de herramientas de alto nivel entre las que se incluyen Spark SQL (para el procesamiento de datos estructurados basada en SQL), MLlib para implementar machine learning, GraphX para el procesamiento de grafos y Spark Streaming. Puede ser descargado de su sitio oficial https://spark.apache.org/downloads.html, es multiplataforma.