

INFORME TÉCNICO FINAL DE PROYECTO

Formato: IFC-001

Formato aprobado por el Comité de Proyectos I+D+i, 13 de junio del 2011

1. DATOS GENERALES

TÍTULO DEL PROYECTO			
<i>Metodología de visualización interactiva y análisis de información en Big Data</i>			
CODIGO DEL PROYECTO		AREA DEL CONOCIMIENTO	
<i>N/A</i>		<i>Tecnologías de la Información y conectividad</i>	
FECHA DE INICIO			
<i>01-04-2016</i>		<i>31-03-2017</i>	
CONVENIO		PLAZO (meses)	
<i>N/A</i>		<i>12 meses</i>	
INSTITUCION EJECUTORA			
<i>CARRERA DE INGENIERÍA EN ELECTRÓNICA Y REDES DE COMUNICACIÓN, FACULTAD DE INGENIERÍA EN CIENCIAS APLICADAS, UNIVERSIDAD TÉCNICA DEL NORTE.</i>			
NOMBRE DIRECTOR DEL PROYECTO			
<i>Ing. Diego Hernán Peluffo Ordóñez, PhD. Investigador 1.</i>			
DIRECCION Y CONTACTOS			
<i>Av. 17 de Julio 521 y Gral. José María Córdova Edificio No. 5, FICA, CARRERA DE INGENIERÍA EN ELECTRÓNICA Y REDES DE COMUNICACIÓN. 062997800 Ext 7506</i>			
COLABORADORES			
Nombre	Función	Teléfono	E-mail
Ana Cristina Umaquina Criollo	Investigador 2	0999631649	acumaquina@utn.edu.ec
Luis Edilberto Suárez Zambrano	Investigador 3	0994517263	lesuarez@utn.edu.ec
Stefany Cristina Flores Armas	Investigador 4	0984513352	scflores@utn.edu.ec
Hernán Mauricio Domínguez Limaico	Investigador 5	0984440136	hmdominguez@utn.edu.ec
Omar Ricardo Oña Rocha	Investigador 6	0996387474	oronia@utn.edu.ec
MONTO TOTAL APOORTE SENESCYT (A)		<i>[Corresponde a todos los desembolsos realizados por SENESCYT durante el proyecto]</i>	
MONTO APOORTE BENEFICIARIA (B)		14710,00 USD	

MONTO TOTAL PROYECTO (A + B)	14710,00 USD
FECHA DE ENTREGA DEL INFORME	
07-04-2017	

2. CONTENIDO

INDICE	
Sección	Página
Resumen	2
Palabras Claves	4
Abstract	4
Glosario de Términos	6
Objetivo General	6
Objetivos Específicos	6
Descripción	7
Productos	18
Impactos	20
Transferencia de Resultados	20
Conclusiones y Recomendaciones	21
Bibliografía	22
Anexos	25

RESUMEN
<p>El termino Big Data es un nuevo concepto que emerge en respuesta al gran volumen de datos que ha alcanzado dimensiones inmanejables, es decir, que supera la capacidad del software habitual para capturar, gestionar y procesar datos en un tiempo razonable, además de las limitaciones de la infraestructura pre-existente. Por lo tanto, Big Data es cualquier atributo que represente un reto a las limitantes de capacidad de los sistemas o las necesidades del estado, la empresa, la educación, la economía, la optimización del tráfico en telecomunicaciones, la salud, el tráfico vehicular urbano, o cualquier otro campo susceptible de generar grandes cantidades de información y, por tanto, conocimiento. Big Data no sólo se refiere al tamaño de los datos, también se refiere al</p>

acelerado ritmo al que se generan datos y la velocidad con que se crean las diversas fuentes que los producen, la cuales pueden ser independientes y tienen la capacidad potencial de interactuar entre sí, propiciando composiciones inconsistentes e impredecibles. En otras palabras, la combinación de pequeñas bases de datos, con algún tipo de relación, puede generar una gigante base de datos.

Sin embargo la capacidad de interpretar, analizar, sintetizar y gestionar dichos volúmenes de datos aumenta a pasos mucho más lentos, las herramientas tradicionales como hojas de cálculo, estadística descriptiva convencional, diagramas o gráficos planos, estáticos y convencionales ya no son suficientes, y es justamente en este punto donde las herramientas deben evolucionar de forma paralela al crecimiento de los datos y estar acordes con los requerimientos de los expertos, por ello se hacen necesarias técnicas emergentes que utilicen las potencialidades de los recursos computacionales y los integre de forma sinérgica a las cualidades de análisis e interpretación humana. Dicha integración, se puede interpretar como una integración de la inteligencia artificial y la inteligencia natural como un equipo idóneo para descubrir conocimiento. Lo anterior es posible aplicando áreas de análisis robusto de datos y extracción eficaz de información tales como la minería de datos y la visualización inteligente de la información.

No obstante, el uso de métodos de minería de datos y técnicas de visualización por separado conlleva a la necesidad de expertos para la interpretación de resultados, siendo esto algo negativo porque involucra un incremento en tiempo, costos y trabajo para llegar a la etapa última del procesamiento de datos, es decir, determinar la información útil y el conocimiento inmersos en los datos analizados. La integración de los métodos de minería y las técnicas de visualización es una necesidad latente; aunque recientemente se ha desarrollado algunos métodos y herramientas de integración, el diseño de un sistema de análisis visual que se adapte adecuadamente a las necesidades y requerimientos de un usuario particular es aún un problema abierto. Una de las formas de hacer una visualización inteligible de grandes volúmenes de datos es reducir la dimensión de los mismos de tal forma que se obtengan representaciones fácilmente interpretables por el ser humano. Sin embargo, el potencial de los métodos de reducción de dimensión ha sido relegado a tareas pre-procesado, descartando la posibilidad de realizar la integración con una reducción interactiva junto con elementos de la analítica visual.

Con el desarrollo de este proyecto, se planea diseñar una metodología de análisis visual de información en Big Data usando principios de visualización e interactividad en conjunto con técnicas de reducción de dimensión. La premisa de esta investigación es la posibilidad de realizar la interacción con el usuario a partir de la inclusión del mismo en la selección y combinación de

métodos de reducción de dimensión. Lo anterior supone diseñar no sólo el modelo y la interfaz de interacción, sino la formulación de un método generalizado de reducción de dimensión que permita realizar intuitivamente la selección y combinación de métodos. Un aspecto transversal de todas las etapas del diseño de la metodología de análisis visual es el costo computacional que obstaculiza la meta de que la metodología sea realmente interactiva, es decir, en tiempo real. Dicho esto, en este proyecto, será determinante realizar todos los diseños e implementaciones de prueba en entornos de bajo costo computacional.

El proyecto se presenta en modalidad de investigación básico, ya que corresponde a la etapa del diseño metodológica de la implementación de una herramienta de análisis y visualización de datos. Dicho diseño servirá de insumo para realizar futuras investigaciones aplicadas y de desarrollo tecnológico en donde se divulgue el uso de la herramienta. Además, se espera que esta iniciativa permita la conformación de una red de investigadores en el área emergente y de alto interés actual de Big Data.

PALABRAS CLAVES

Big Data, reducción de dimensión, machine learning, técnicas de reducción de dimensión, minería de datos, dimensionalidad, procesos KDD, open data, métodos espectrales, business intelligence, analítica de datos.

ABSTRACT

The term Big Data is a new concept that emerges in response to the large volume of data, which has reached unmanageable dimensions -exceeding the ability of the usual software to capture, manage and process data in a reasonable time, in addition to the limitations of the pre-existing infrastructure. In other words, Big Data is any attribute representing a challenge to the capacity constraints of the systems or the needs of the state, business, education, economy, traffic optimization in telecommunications, health, urban vehicular traffic. As well, any other field capable of generating large amounts of information, which, once processed, becomes knowledge. Big Data not only refers to the size of the data, but also refers to the accelerated rate at which data is generated and the speed with which the various sources that produce it are created, which can be independent and have the potential to interact among them. Indeed, the combination of small databases, with some kind of relationship, can generate a giant database.

However, the ability to interpret, analyze, synthesize and manage such data volumes increases at

much slower steps, traditional tools such as spreadsheets, conventional descriptive statistics, diagrams or flat, static and conventional charts are no longer sufficient, and at this point where the tools must evolve in parallel with the growth of the data and be in accordance with the requirements of the experts, therefore it becomes necessary emerging techniques that use the potential of the computational resources and integrate them in a synergistic way to the Qualities of human analysis and interpretation. Such integration can be interpreted as an integration of artificial intelligence and natural intelligence as an ideal team to discover knowledge. This is possible by applying areas of robust data analysis and efficient extraction of information such as data mining and intelligent visualization of information.

Nonetheless, the use of data mining methods and visualization techniques separately leads to the need for experts to interpret results, this being a negative thing because it involves an increase in time, costs and work to reach the final stage of data processing procedures. The integration of data mining methods and visualization techniques is a latent necessity. Despite some integration methods and tools have recently been developed, the design of a visual analysis system that adequately adapts to the needs and requirements of a particular user is still an open issue. One way to reach an intelligible visualization of large volumes of data is via dimensionality reduction aimed at reaching representations being easily interpretable by the human being. The potential of dimensional reduction methods has been relegated to pre-processed tasks, ruling out the possibility of performing the integration with an interactive reduction along with elements of visual analytics.

The aim of this project is to design a methodology of visual analysis of information in Big Data using principles of visualization and interactivity in conjunction with techniques of reduction of dimension. The premise of this research is the possibility of performing the interaction with the user during the selection and combination of methods of reduction of dimension. The aforementioned involves designing not only the interaction model and interface, but also the formulation of a generalized method of dimension reduction that allows intuitive selection and combination of methods. A transversal aspect of all the stages for designing the visual analysis methodology is the computational cost that hinders the goal of the methodology being really interactive, that is, in real time. That said, in this project, it will be crucial to perform all test designs and implementations in low computational computing environments.

The project is presented in a basic research modality, since it corresponds to the stage of the methodological design of the implementation of a data analysis and visualization tool. This design will serve as input for future applied research and technological development where the use of the

tool is disclosed. In addition, it is expected that this initiative will allow the formation of a network of researchers in the emerging and high interest area of Big Data.

GLOSARIO DE TERMINOS

- **BIG DATA:** Llamado también datos masivos, conjuntos de datos tan grandes que no es posible procesarlos con los sistemas convencionales de bases de datos.
- **MINERÍA DE DATOS:** Conjunto de técnicas usadas para el análisis de conjuntos de datos y encontrar relaciones, patrones que sean entendibles y útiles para el usuario.
- **KDD:** Acrónimo en inglés de Knowledge Discovery Data Bases, es el descubrimiento de conocimiento en bases de datos, es decir, es un proceso para encontrar patrones entre los datos que sean de utilidad para el usuario.
- **OPEN DATA:** Es todo conjunto de datos que ha sido distribuido libremente a través de la red, sin restricciones de uso, ni derechos de autor.
- **REDUCCIÓN DE DIMENSIÓN:** Proceso de reducción del número de variables aleatorias que se trate, y se puede dividir en selección de función y extracción de función.
- **APRENDIZAJE AUTOMÁTICO:** Subcampo de las ciencias de la computación y una rama de la inteligencia artificial cuyo objetivo es desarrollar técnicas que permitan a las computadoras aprender. Se trata de crear programas capaces de generalizar comportamientos a partir de una información suministrada en forma de ejemplos. Es, por lo tanto, un proceso de inducción del conocimiento.
- **VISUALIZACIÓN DE DATOS:** Proceso de búsqueda, interpretación, contrastación y comparación de datos que permite un conocimiento en profundidad y detalle de los mismos de tal forma que se transformen en información comprensible para el usuario.
- **PRINCIPAL COMPONENT ANALYSIS:** Es un método espectral de análisis multivariado, basado en correlaciones entre variables, mediante un proceso estocástico que satisface la reducción de dimensiones, realizando una transformación lineal de las variables iniciales, para proyectar vectores propios ortonormales denominados componentes principales.

- **LINEAR DISCRIMINANT ANALYSIS:** Análisis discriminante lineal bajo el criterio de Fisher, este método busca las muestras que sigan patrones de distribución gaussianas para un buen funcionamiento.

OBJETIVO GENERAL

Desarrollar una metodología de visualización interactiva y eficaz de información en Big Data, usando un modelo de interacción y técnicas de reducción de dimensión, que presente un buen compromiso entre desempeño en la representación de los datos y costo computacional.

OBJETIVOS ESPECÍFICOS

Objetivo Específico 1: Seleccionar técnicas representativas de reducción de dimensión, teniendo en cuenta criterios de desempeño y costo computacional, con el fin de determinar las más adecuadas para aplicaciones de Big Data.

Objetivo Específico 2: Proponer un método generalizado de reducción de dimensión que permita representar diversos métodos existentes y que presente bajo costo computacional.

Objetivo Específico 3: Formular un modelo matemático de interacción usuario-computador que habilite al usuario para realizar la combinación y/o selección de métodos reducción de dimensión a través de la visualización interactiva de los datos.

Objetivo Específico 4: Diseñar un esquema de visualización intuitivo e inteligible de datos reducidos de Big Data empleando técnicas modernas de analítica visual y elementos de la percepción humana.

Objetivo Específico 5: Desarrollar un sistema completo para el análisis visual de bases de datos de Big Data, usando un modelo de interacción y técnicas de reducción de dimensión, que permita la visualización eficaz e interactiva de datos.

DESCRIPCIÓN

En concordancia con el Objetivo 11, Política y Estrategia #3 literales c) y n), del Plan Nacional del Buen Vivir 2013-2017 del Ecuador, el proyecto tiene lugar puesto que representa una contribución a la calidad, la seguridad y la cobertura en la prestación de servicios públicos usando las telecomunicaciones

para garantizar la soberanía y la seguridad en la gestión de la información en diferentes procesos dentro de las instituciones y las empresas [57].

En la actualidad, existe la necesidad de integrar los sistemas de descubrimiento de conocimiento en grandes conjuntos de datos (DCBD) con el fin de identificar la información más importante que se encuentra inmersa en una serie de variables contenidas en grandes volúmenes de información, de tal forma que sea fácilmente interpretable a través de una visualización clara y coherente de estos resultados del análisis [6]. Las herramientas de DCBD forman parte de un área activa de investigación que se encuentra en constante evolución [7, 8], por tanto, los resultados esperados de este proyecto de trabajo de grado representan un importante aporte en el área de visualización y minería de datos garantizando eficiencia y aplicabilidad. Las diferentes herramientas que hoy en día se han implementado utilizan diferentes técnicas que permiten la clasificación de datos teniendo en cuenta el alcance y las etapas del DCBD. Algunas se enfocan únicamente en la minería de datos teniendo por objetivo la clasificación de información sin tener muy cuenta la representación visual de los datos ni los requerimientos del usuario que debe realizarla [6]. Otras herramientas existentes integran la visualización con los sistemas de minería de datos con el fin de realizar una aplicación completa sin embargo exigen un grado de conocimiento teórico para poder interpretar los resultados finales.

Una técnica de minería de datos visual no es justamente una técnica de visualización, siendo enfocada a explotar los datos en algunas fases de un proceso de minería de datos, pero un algoritmo de minería de datos con análisis visual podría desempeñar un mejor papel [9]. La implementación de una interfaz interactiva basada en la combinación de métodos de reducción de dimensión desarrollada dentro de un algoritmo de análisis visual es una nueva metodología dentro de la minería de datos, y por consiguiente dentro del DCBD, que permitirá a un usuario interactuar con los métodos, sin que necesariamente se tenga conocimiento previo sobre estos, y obtener resultados deseados. Así se obtendría una nueva forma de realizar la visualización de datos a partir de la reducción de dimensión, orientada a la integración interactiva y eficiente de estas dos formas de conocimiento, es decir: Trabajo conjunto entre métodos de minería de datos y análisis visual.

Lo que se busca con el desarrollo de este proyecto es realizar una metodología completa de visualización de información de Big Data de forma que exista una interacción entre usuario y computador. Dicha interacción debe existir incluso sin necesidad de que el usuario conozca el algoritmo de funcionamiento, es decir que el modelo de interacción es de uso intuitivo y amigable con el usuario. El proyecto será apoyado por el desarrollo de tesis doctorales, de maestría y de pregrado. La

metodología propuesta servirá de base para futuras investigaciones encaminadas a la implementación de una herramienta completa de DCDB que integrará el análisis visual y los procesos de minería de datos. Adicionalmente, cabe anotar que los investigadores participantes del proyecto cuentan con asesorías de expertos nacionales y extranjeros.

EL PROBLEMA

El descubrimiento de conocimiento en base de datos (DCBD) es en sí un proceso con el cual se extraen patrones a partir de conjuntos de datos, este proceso es automático y combina descubrimiento y análisis [26]. El DCBD ya ha sido ampliamente explorado y desarrollado, y tiene una gran diversidad de aplicaciones como la determinación de perfiles de clientes fraudulentos [2,3], descubrir una relación implícita que exista entre síntomas y enfermedades [4,8], análisis de mercado, ventas y soluciones a clientes [9], análisis del genoma de la levadura [20], aplicaciones bioinformáticas [21], entre otras aplicaciones, es por esto que se requiere que las técnicas de DCBD se sigan explorando y desarrollando. Muchos investigadores han reconocido la necesidad de integrar los sistemas de DCBD en la exploración de grandes volúmenes de datos (BigData, Massive data), dando lugar a nuevas áreas de investigación [10, 11, 19, 21].

En la actualidad existen herramientas que implementan técnicas que implican el pre procesamiento, el uso de métodos de minería de datos y/o visualización [12, 18, 21, 22], estas herramientas realizan el proceso de DCBD pero presentan diversos problemas entre los que se encuentra una respuesta aceptable pero con un gasto computacional alto y además con un bajo rendimiento en velocidad, o una respuesta rápida pero con resultados muy difícilmente interpretables, es decir, para poder analizar los resultados es necesario contar con un conocimiento previo o tener experiencia en dicha interpretación [13, 14], también se considera que estas herramientas no mantienen una apropiada interactividad con el usuario debido a que no permite adecuar los parámetros existentes en los grandes volúmenes de datos y no toman a consideración las necesidades del usuario [19].

Las herramientas de visualización de grandes volúmenes de datos se encargan de aplicar ya sea los métodos de minería de datos para la extracción de patrones en formas de reglas o funciones, o métodos de reducción de dimensión orientados a la visualización, pero sea cual sea el caso, realizar la visualización de dichas dimensiones reducidas puede llegar a ser algo abstracto [14, 15, 16]. Por tanto, se aprecia que existe la necesidad de que el usuario cuente con una herramienta interactiva que permita

manipular los métodos de DCBD acorde a sus necesidades para obtener los resultados deseados y que éstos sean fácilmente interpretables sin que se tenga un conocimiento previo o básico de tales métodos [19, 22]. De aquí que la representación visual cumple un papel muy importante dentro del procesamiento de datos cuando se fusiona y complementa con los procesos de minería de datos [17, 19].

METODOLOGÍA

Objetivo específico 1:

Seleccionar técnicas representativas de reducción de dimensión, teniendo en cuenta criterios de desempeño y costo computacional, con el fin de determinar las más adecuadas para aplicaciones de Big Data.

Actividad 1.1.

Revisión bibliográfica exhaustiva y elaboración de estado del arte en técnicas de reducción de dimensión.

Investigador 1
Investigador 5

METODOLOGÍA:

- ✓ Búsqueda de artículos en bases de datos científicas tales como IEEE, SCIENCE DIRECT, SPRINGER y sitios web especializados relacionados con técnicas de reducción de dimensión de tipo espectral.

TAREAS:

- ✓ Descarga y revisión de artículos
- ✓ Selección de información válida, asociada a la temática
- ✓ Escritura del estado del arte en técnicas de reducción de dimensión.
- ✓ Organización estratégica, conceptual y cronológica del estado del arte.

EVIDENCIA:

- ✓ Ver **Anexo A**: Estado del arte en Reducción de dimensión.

<p>Actividad 1.2. Establecer y comparar las técnicas de reducción de dimensión.</p>	<p>Investigador 1 Investigador 3 Investigador 5 Investigador 6</p>
<p>METODOLOGÍA:</p> <ul style="list-style-type: none"> ✓ Implementación en lenguaje de programación las técnicas de reducción de dimensión modernas. ✓ Diseño y ejecución de una metodología de comparación de técnicas de reducción de dimensión. 	
<p>TAREAS:</p> <ul style="list-style-type: none"> ✓ Programación en Matlab y Processing de técnicas de reducción de dimensión. ✓ Desarrollo de medidas cuantitativas para la comparación de métodos espectrales de reducción de dimensión. 	
<p>EVIDENCIA:</p> <ul style="list-style-type: none"> ✓ Ver Anexo B: Estudio comparativo de métodos espectrales para reducción de la dimensionalidad: LDA versus PCA. 	
<p>Objetivo específico 2: Proponer un método generalizado de reducción de dimensión que permita representar diversos métodos existentes y que presente bajo costo computacional.</p>	
<p>Actividad 2.1. Revisión bibliográfica exhaustiva y elaboración de estado del arte en integración y métodos generalizados de reducción de dimensión para Big Data.</p>	<p>Investigador 1 Investigador 5</p>
<p>METODOLOGÍA:</p> <ul style="list-style-type: none"> ✓ Búsqueda de artículos en bases de datos científicas tales como IEEE, SCIENCE DIRECT, 	

<p>SPRINGER y sitios web especializados relacionados con métodos generalizados de reducción de dimensión de tipo espectral y técnicas de integración con el usuario.</p>	
<p>TAREAS:</p> <ul style="list-style-type: none"> ✓ Descarga y revisión de artículos ✓ Selección de información válida, asociada a la temática ✓ Escritura del estado del arte en técnicas de reducción de dimensión. ✓ Organización estratégica, conceptual y cronológica del estado del arte. 	
<p>EVIDENCIA:</p> <ul style="list-style-type: none"> ✓ Ver Anexo C: estado del arte en integración y métodos generalizados de reducción de dimensión. ✓ Ver Anexo I: Publicaciones. Artículo “<i>On the Relationship Between Dimensionality Reduction and Spectral Clustering from a Kernel Viewpoint</i>”, el cual describe el esquema generalizado o unificado de métodos de reducción que corresponde al <u>Resultado Esperado 5.3</u> y es un producto de transferencia de resultados. 	
<p>Actividad 2.2. Diseño e implementación en lenguaje de programación del método generalizado de reducción de dimensión.</p>	<p>Investigador 1 Investigador 3 Investigador 5</p>
<p>METODOLOGÍA:</p> <ul style="list-style-type: none"> ✓ Definición de los lenguajes y entornos de programación para desarrollar el método generalizado. 	
<p>TAREAS:</p> <ul style="list-style-type: none"> ✓ Empleo y desarrollo toolboxes especializados para reducción de dimensión y visualización en MatLab y Processing. 	
<p>EVIDENCIA:</p> <ul style="list-style-type: none"> ✓ Ver Anexo D: Diagrama de flujo del esquema propuesto. ✓ Ver Sitio web del proyecto: https://sites.google.com/view/metodologiabigdata/software?authuser=0, opción “Interfaces de Visualización”. <u>Resultado Esperado 5.3</u>. 	

Objetivo específico 3: Formular un modelo matemático de interacción usuario-computador que habilite al usuario para realizar la combinación y/o selección de métodos reducción de dimensión a través de la visualización interactiva de los datos.	
Actividad 3.1. Revisión bibliográfica exhaustiva y elaboración de estado del arte en modelos de interactivos de combinación y selección de métodos de reducción de dimensión.	Investigador 1 Investigador 6
METODOLOGÍA: <ul style="list-style-type: none"> ✓ Búsqueda de artículos en bases de datos científicas tales como IEEE, SCIENCE DIRECT, SPRINGER y sitios web especializados relacionados con modelos de interacción para incluir al usuario en la selección o combinación de métodos de reducción de dimensión. 	
TAREAS: <ul style="list-style-type: none"> ✓ Descarga y revisión de artículos ✓ Selección de información válida, asociada a la temática ✓ Escritura del estado del arte en técnicas de reducción de dimensión. ✓ Organización estratégica, conceptual y cronológica del estado del arte. 	
EVIDENCIA: <ul style="list-style-type: none"> ✓ Ver Anexo E: Modelo de interacción para el análisis visual de datos en Big Data usando reducción de dimensión interactiva <u>Resultado Esperado 5.1.</u> 	
Actividad 3.2. Formulación e implementación en lenguaje de programación del modelo propuesto.	Investigador 1 Investigador 3 Investigador 6
METODOLOGÍA: <ul style="list-style-type: none"> ✓ Definición de los lenguajes y entornos de programación para desarrollar el modelo interactivo para la selección y/o combinación de métodos de reducción de dimensión. 	
TAREAS: <ul style="list-style-type: none"> ✓ Empleo y desarrollo toolboxes especializados para reducción de dimensión y visualización 	

interactiva en MatLab y Processing.	
EVIDENCIA: ✓ Ver Sitio web del proyecto: https://sites.google.com/view/metodologiabigdata/software?authuser=0 , opción “Interfaces de Visualización”. <u>Resultado Esperado 5.1.</u>	
Actividad 3.3 Envío de artículo a revista internacional y/o ponencia nacional o internacional.	Investigador 1 Investigador 2 Investigador 3 Investigador 4 Investigador 5 Investigador 6
METODOLOGÍA: ✓ Realización de un escrito para reportar el modelo interactivo propuesto y los resultados alcanzados.	
TAREAS: ✓ Escribir y enviar un artículo a un evento internacional para divulgar el modelo propuesto.	
EVIDENCIA: ✓ Ver Anexo I: Publicaciones. Artículo “ <i>Interactive visualization methodology of high-dimensional data with a color-based model for dimensionality reduction</i> ”, el cual describe el modelo interactivo basado en colores que corresponde al <u>Resultado Esperado 5.1</u> y es un producto de transferencia de resultados .	
Objetivo específico 4: Diseñar un esquema de visualización intuitivo e inteligible de datos reducidos de Big Data empleando técnicas modernas de analítica visual y elementos de la percepción humana.	
Actividad 4.1. Revisión bibliográfica exhaustiva y elaboración de estado del arte en visualización de datos y percepción humana.	Investigador 1 Investigador 2 Investigador 4
METODOLOGÍA: ✓ Búsqueda de artículos en bases de datos científicas tales como IEEE, SCIENCE DIRECT, SPRINGER y sitios web especializados relacionados con métodos de visualización de	

datos basados en principios de percepción humana (color, forma, estructura).	
TAREAS: <ul style="list-style-type: none"> ✓ Descarga y revisión de artículos ✓ Selección de información válida, asociada a la temática ✓ Escritura del estado del arte en técnicas de reducción de dimensión. ✓ Organización estratégica, conceptual y cronológica del estado del arte. 	
EVIDENCIA: <ul style="list-style-type: none"> ✓ Ver Anexo G: Cartilla tutorial sobre técnicas y pautas de análisis visual basado en forma y color. 	
Actividad 4.2. Diseñar el esquema de visualización usando conceptos de la teoría de la percepción humana.	Investigador 1 Investigador 4
METODOLOGÍA: <ul style="list-style-type: none"> ✓ Elaboración del esquema de visualización con principios de percepción humana. 	
TAREAS: <ul style="list-style-type: none"> ✓ Implementar la interfaz de visualización. 	
EVIDENCIA: <ul style="list-style-type: none"> ✓ Ver Sitio web del proyecto: https://sites.google.com/view/metodologiabigdata/software?authuser=0, opción “Interfaces de Visualización”. <u>Resultado Esperado 5.2.</u> 	
Actividad 4.3. Implementación del esquema de visualización en lenguajes de programación.	Investigador 1 Investigador 2 Investigador 3
METODOLOGÍA: <ul style="list-style-type: none"> ✓ Definición de los lenguajes y entornos de programación para desarrollar el esquema de visualización. 	
TAREAS:	

<p>✓ Empleo y desarrollo toolboxes especializados para visualización de datos en MatLab y Processing.</p>	
<p>EVIDENCIA:</p> <p>✓ Ver Sitio web del proyecto: https://sites.google.com/view/metodologiabigdata/software?authuser=0, opción “Interfaces de Visualización”. <u>Resultado Esperado 5.2.</u></p>	
<p>Actividad 4.4. Envío de artículo a revista internacional y/o ponencia nacional o internacional.</p>	<p>Investigador 1 Investigador 2 Investigador 3 Investigador 4 Investigador 5 Investigador 6</p>
<p>METODOLOGÍA:</p> <p>✓ Realización de un escrito para reportar el esquema de visualización propuesto y los resultados alcanzados.</p>	
<p>TAREAS:</p> <p>✓ Escribir y enviar un artículo a un evento internacional para divulgar el modelo propuesto.</p>	
<p>EVIDENCIA:</p> <p>✓ Ver Anexo I: Publicaciones. Artículo “<i>PROPUESTA DE ANÁLISIS VISUAL DE DATOS EN BIG DATA USANDO REDUCCIÓN DE DIMENSIÓN INTERACTIVA</i>”, el cual describe el esquema visualización que corresponde al <u>Resultado Esperado 5.1</u> y es un producto de transferencia de resultados.</p>	
<p>Objetivo específico 5:</p> <p>Desarrollar un sistema completo para el análisis visual de bases de datos de Big Data, usando un modelo de interacción y técnicas de reducción de dimensión, que permita la visualización eficaz e interactiva de datos.</p>	
<p>Actividad 5.1. Desarrollo e implementación del sistema completo de análisis visual.</p>	<p>Investigador 1 Investigador 2</p>

		Investigador 3 Investigador 4 Investigador 5 Investigador 6
METODOLOGÍA: ✓ Definición de los lenguajes y entornos de programación para desarrollar un sistema completo de visualización con interacción síncrona y parámetros subóptimos.		
TAREAS: ✓ Empleo y desarrollo toolboxes especializados para visualización de datos e interacción síncrona en MatLab y Processing.		
EVIDENCIA: ✓ Ver Sitio web del proyecto: https://sites.google.com/view/metodologiabigdata/software?authuser=0 , opción “Interfaces de Visualización”. <u>Resultado Esperado 5.4 y 5.6.</u>		
Actividad 5.2. Realización de cursos de capacitación en Big Data y escritura de artículos.		Investigador 1 Investigador 2 Investigador 3 Investigador 4 Investigador 5 Investigador 6
METODOLOGÍA: ✓ Desarrollo de cursos de capacitación para personal académico de la Universidad Técnica del Norte en temas de visualización de información de Big Data y escritura de artículos científicos en ingeniería.		
TAREAS: ✓ Organización, gestión y orientación de cursos de visualización de información de Big Data y Escritura de artículos científicos en ingeniería para la comunidad académica de la Universidad Técnica del Norte.		

EVIDENCIA:

- ✓ Ver **Anexo H:** Cursos dictados.

Actividad 5.3. Dirección del proyecto, asesoría a investigadores participantes y formación de recurso humano.

Investigador 1

METODOLOGÍA:

- ✓ Asesoría y dirección del proyecto en general.
- ✓ Formación de investigadores a nivel de pregrado, maestría y doctorado.

TAREAS:

- ✓ Realización de reuniones.
- ✓ Asesorías a estudiantes e investigadores.

EVIDENCIA:

- ✓ Ver Anexo J: Actas de reuniones.
- ✓ Ver Anexo K: Formación de recurso humano.
- ✓ Ver Anexo I: Publicaciones.

PRODUCTOS

RESULTADOS Y DISCUSIÓN

Como resultados del Proyecto de investigación se presentan los productos que se relaciona en la siguiente tabla:

META	DESCRIPCIÓN	META ALCANZADA
Envío de artículos científicos a revistas internacionales de alto impacto	Envío de 2 (dos) artículos a revistas, internacionales y reconocidas, relacionadas con el área de inteligencia artificial y <i>machine learning</i> . Entre las revistas candidatas se contempla: Lecture Notes in Computer Science y Knowledge-based Systems de la editorial Springer.	Se logró la publicación de 2 (dos) artículos científicos en las revistas: ADCAIJ, enfoque UTE. Adicionalmente se logró el envío de 1 (uno) artículo a la revista ASTES Journal como consta en las evidencias del Anexo I .
Ponencias nacionales	Participación en 2 (dos) ponencias y conferencias nacionales como ETCM2016, INCISCOS.	Se logró la aceptación y disertación de 7 (siete) ponencias en eventos nacionales tales como: 1 en CONGRESO INTERNACIONAL DE

		INGENIERÍAS UPEC 2016, 2 en ETCM 2016, 1 en las JORNADAS INTERNACIONALES FICA 2016, 2 en INCISCOS UTE 2016, CIITC 2016, como consta en las evidencias del Anexo I .
Ponencias internacionales	Participación en 2 (dos) ponencias y conferencias internacionales. Se contempla las siguientes conferencias: 2016 IEEE LA-CCI (Latin American Conference on Computational Intelligence) y STSIVA 2016 (XXI Symposium On Image, Signal Processing And Artificial Vision).	Se logró la aceptación y disertación de 5 (cinco) ponencias en eventos internacionales tales como: DCAI'16, IEEE LA-CCI 2016, STSIVA, CIARP 2016, IWINAC 2017. como consta en las evidencias del Anexo I .
Curso de escritura de artículos científicos en ingeniería	Impartición 1 (uno) curso sobre pautas generales de forma y fondo para la escritura de artículos científicos en ingeniería. Dirigido por investigadores del proyecto y profesores invitados. Duración: 40 horas. Dirigido a las facultades de ingeniería de la Universidad Técnica del Norte.	Como consta en el Anexo H , se impartió el curso que se tenía como meta Curso: Escritura de artículos de investigación en Ingeniería, junio 2016 .
Curso básico de análisis y visualización de información en Big Data	Impartición de 1 (uno) curso sobre herramientas y métodos de análisis y visualización de información en datos de alta dimensión. Dirigido por investigadores del proyecto y profesores invitados. Duración: 40 horas. Dirigido a las facultades de ingeniería de la Universidad Técnica del Norte.	Como consta en el Anexo H , se impartió el curso que se tenía como meta Curso: Análisis Visual de la información en Big Data, en marzo 2017 .

Adicionalmente, como resultados del proyecto se han logrado los siguientes artículos de libro:

1. En el evento TECNOLOGÍAS APLICADAS A LA INGENIERÍA, EDITORIAL UTN con ISBN: 978-9942-984-17-3.
2. GENERANDO CIENCIA MEMORIAS DE LAS I JORNADAS INTERNACIONALES DE INVESTIGACIÓN CIENTÍFICA, EDITORIAL UTN, con ISBN: 978-9942-984-05-0.

Por otro lado, se han conseguido otros resultados como parte del proyecto.

1. Se logró la culminación de **2 (dos)** Cursos, por parte de dos integrantes del grupo del proyecto de investigación en las temáticas: como consta en el **Anexo H**.

- a. IntroDatax: Big Data for Smart Cities, IEEE, Julio 2016, Ing. Luis Suárez Zambrano
 - b. Taller práctico: Big Data & Data Analytics, Handytecmobi S.A., noviembre 2016, Ing. Ana Umaquina Criollo.
2. Se logró la impartición de **1 (un)** Curso de Aprendizaje en el aula Universitaria, impartido por la Lcda. Stefany Flores integrante del proyecto.

También se logró varias conferencias nacionales e internacionales en calidad de Conferencista Magistral Invitado, en los siguientes eventos:

- a. I SIMPOSIO INTERNACIONAL DE INGENIERÍA MECATRÓNICA, UNIVERSIDAD MARIANA-PASTO, marzo 2017.
 - b. AUNAR DATA VIS DAY, PASTO-COLOMBIA, agosto 2016.
 - c. I CONGRESO INTERNACIONAL DE C+T+I Y DESARROLLO TERRITORIAL, PASTO-COLOMBIA, octubre 2016.
 - d. JORNADAS INTERNACIONALES DE INVESTIGACIÓN CIENTÍFICA, IBARRA-ECUADOR, UTN 2016, julio 2016.
 - e. DCAI 2016, SEVILLA-ESPAÑA, junio 2016.
3. Finalmente se ha logrado presentar un poster en el Congreso Internacional de Innovación y Transferencia de Conocimiento CIITC 2016, con el título ***“Mejoramiento de las competencias y desarrollo de las capacidades de desempeño de los actores específicos EPS mediante la capacitación y asesoramiento de la utilización de herramientas informáticas y de procesos técnicos en las áreas de Electrónica”***, octubre 2016, ESPE – QUITO-ECUADOR”

Todas estas evidencias constan en el **Anexo I**, de publicaciones.

IMPACTOS

Este proyecto permitió a los investigadores participantes reforzar y desarrollar habilidades para la investigación básica en aspectos tales como: Formulación matemática, gestión de la información, programación optimizada, y escritura científica en inglés. Otros beneficiarios directos del proyecto son los usuarios finales del prototipo de la interfaz, quienes pueden disfrutar de la herramienta y pueden realizar una realimentación sobre su experiencia para posibles mejoras del diseño.

A nivel científico, se beneficia a la comunidad académica en el área y a la comunidad universitaria

en general con informes técnicos y con la generación de nuevo conocimiento evidenciada en artículos de investigación de impacto.

TRANSFERENCIA DE RESULTADOS

Con el desarrollo de este proyecto, se espera lograr por lo menos las siguientes metas de transferencia de resultados:

META	CANTIDAD	DESCRIPCIÓN
Envío de artículos científicos a revistas internacionales de alto impacto	2	Envío de 2 artículos a revistas, internacionales y reconocidas, relacionadas con el área de inteligencia artificial y <i>machine learning</i> . Entre las revistas candidatas se contempla: Lecture Notes in Computer Science y Knowledge-based Systems de la editorial Springer.
Ponencias nacionales	2	Participación en ponencias y conferencias nacionales como ETCM2016, INCISCOS.
Ponencias internacionales	2	Participación en ponencias y conferencias internacionales. Se contempla las siguientes conferencias: 2016 IEEE LA-CCI (Latin American Conference on Computational Intelligence) y STSIVA 2016 (XXI Symposium On Image, Signal Processing And Artificial Vision).
Curso de escritura de artículos científicos en ingeniería	1	Impartición un curso sobre pautas generales de forma y fondo para la escritura de artículos científicos en ingeniería. Dirigido por investigadores del proyecto y profesores invitados. Duración: 40 horas. Dirigido a las facultades de ingeniería de la Universidad Técnica del Norte.
Curso básico de análisis y visualización de información en Big Data	1	Impartición de un curso sobre herramientas y métodos de análisis y visualización de información en datos de alta dimensión. Dirigido por investigadores del proyecto y profesores invitados. Duración: 40 horas. Dirigido a las facultades de ingeniería de la Universidad Técnica del Norte.

CONCLUSIONES Y RECOMENDACIONES

Conclusiones:

Con el desarrollo de este proyecto, considerando los resultados presentados en este informe y las evidencias de los anexos, se concluye lo siguiente:

- ✓ Cuando se requiere del diseño de un sistema de visualización de grandes volúmenes de información de forma inteligible para el ser humano, los métodos de reducción de dimensión son una buena alternativa. En este proyecto, se comprobó que, especialmente, las técnicas espectrales de reducción de dimensión son adecuadas dada su versatilidad, factibilidad de representación en matrices *kernel* y buen desempeño para formar grupos de datos separables y preservar la estructura geométrico y topológica de todo el conjunto de datos en un espacio de menor dimensión. Por tanto, dichas técnicas resultan útiles para el diseño de interfaces de visualización de información en Big Data.
- ✓ Para realizar evaluar técnicas de reducción de dimensión, se requiere de un método generalizado para realizar de forma justa la comparación, selección y/o combinación de técnicas. En este sentido, se propone un método generalizado que permite representar técnicas espectrales de reducción de dimensión. Particularmente, el método propuesto se basa en representaciones kernel y sigue un criterio geométrico y de varianza acumulada.
- ✓ El diseño de una interfaz interactiva de visualización basada en reducción de dimensión puede realizarse a través de un modelo de interacción de selección y/o combinación de técnicas. El modelo desarrollado en este proyecto se basa en una suma ponderada de matrices kernel que representan técnicas específicas de reducción de dimensión.
- ✓ El sentido de interactividad incluido en el esquema propuesto se refiere a la facultad del usuario de modificar de forma intuitiva los factores de ponderación de la suma ponderada de matrices kernel, usando enfoques cromáticos, geodésico, geométrico, basados en ecualizadores y similitudes, los cuales cumplen criterios de percepción humana inteligible.
- ✓ Como se demostró en este proyecto, un sistema completo para el análisis visual de bases de datos de Big Data eficiente puede lograrse a través de un modelo de interacción y técnicas de reducción de dimensión

Recomendaciones:

- ✓ Es necesario promover y aprovechar al máximo dentro de la comunidad universitaria de la

UTN, y de la comunidad académica en general el uso de la herramienta desarrollada en esta investigación, ya que ésta permite el análisis de Big Data enfocados a establecer una herramienta de software y técnicas de visualización, la misma que está disponible en el sitio web del proyecto; brindando así información de calidad y de fácil uso.

- ✓ Es recomendable explorar más técnicas de reducción de dimensión que usen otros criterios (además de espectrales) y que se acoplen adecuadamente a modelos de interacción y enfoques de visualización.

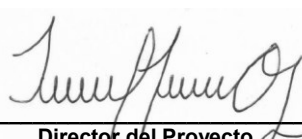
BIBLIOGRAFIA

- [1] E. Bertini y D. Lalanne, «"Surveying the complementary role of automatic data analysis and visualization in knowledge discovery" Proceedings of the ACM SIGKDD Workshop on Visual Analytics and Knowledge Discovery: Integrating Automated Analysis with Interactive Exploration,» ACM, 2009.
- [2] D. H. Peluffo-Ordoñez, J. A. Lee y M. Verleysen, «Short review of dimensionality reduction methods based on stochastic neighbor embedding». Advances in Self-Organizing Maps and Learning Vector Quantization.,» Springer International Publishing, pp. 65-74, 2014.
- [3] I. Borg y P. Groenen, Modern multidimensional scaling: Theory and applications. Springer Science & Business Media, 2005.
- [4] W. Dai y H. Peng, «"Research on Personalized Behaviors Recommendation System Based on Cloud Computing.",» TELKOMNIKA Indonesian Journal of Electrical Engineering 12.2, pp. 1480-1486, 2013.
- [5] M. Ward, G. Grinstein y D. Keim, «Interactive data visualization: foundations, techniques, and applications,» AK Peters, Ltd., 2010.
- [6] P. C. Wong, «Visual Data Mining,» de Computer Graphics and Applications., 1999, pp. 20-21.
- [7] Q. Yang y X. Wu, «10 challenging problems in data mining research,» International of Information Technology & Decision Making, 5(04), pp. 597-604, 2006.
- [8] H. Geppert, M. Vogt y J. Bajorath, «Current trends in ligand-based virtual screening: molecular representations, data mining methods, new application areas, and performance evaluation,» Journal of chemical information and modeling, 50(2), pp. 205-216, 2010.
- [9] J. C. Riquelme, R. Ruiz y K. Gilbert, «Inteligencia Artificial,» de Minería de Datos: Conceptos y tendencias, vol. 10, 2006.
- [10] S. J. Vallejos, Minería de Datos, Universidad Nacional del Nordeste, Argentina, 2006.
- [11] D. A. Keim, «Visual Database Exploration Techniques,» de Visual Techniques for Exploring Databases, 1997.
- [12] D. Keim, F. Mansmann, J. Schneidewind y H. Ziegler, «Challenges in Visual Data Analysis,» de Proceedings of Information Visualization, 2006, pp. 9-16.
- [13] C. Ahlberg y E. Wistrand, «An Information Visualization and Exploration Environment,» Int'l Symp, Information Visualization, pp. 66-73, 1995.
- [14] A. Kerren, A. Ebert y J. Meyer, Human-centered visualization environments, 2006.
- [15] C. P. Lopez, Minería de datos: técnicas y herramientas, Paraninfo, 2007.
- [16] M. Tascón, «Introducción: Big Data. Pasado, presente y futuro,» de Telos: Cuadernos de comunicación e innovación, 2013, pp. 47-50.
- [17] D. Pimentel, M. Cataldi y G. Muñiz, «De la visualización a la sensorización de información,» Blucher Design Proceedings, pp. 129-133, 2013.
- [18] J. C. Alvarado-Perez y H. Bolaños Ramires, Descubrimiento de conocimiento en bases de datos: La perspectiva de la visualización inteligente de la información, 2014.
- [19] D. A. Keim y H. P. Kriegel, «Visualization Techniques for Mining Large Databases: A compraison,»

- IEEE Trans. Knowledge and Data Eng, 8(6), pp. 923-936, 1996.
- [20] W. S. Cleveland, «Visualizing Data,» Horbart Press, 1993.
- [21] A. Inselberg y B. Dimsdale, «Parallel Coordinates: A Tool for Visualizing Multidimensional Geometry,» IEEE Visualization '90, pp. 361-375, 1990.
- [22] A. Inselberg, «The Plane with Parallel Coordinates,» de The Visual Computer, 1985, pp. 69-91.
- [23] A. Inselberg, «Multidimensional Detective,» Information Visualization (InfoVis '97), pp. 100-107, 1997.
- [24] P. E. Hoffman, Table Visualizations: A Formal model and Its Applications, University of Massachusetts.
- [25] D. A. Keim y H. P. Kriegel, «VisDB: Database Exploration Using Multidimensional Visualization,» de Computer Graphics and Applications, 1994, pp. 40-49.
- [26] C. G. Beshers y S. K. Feiner, «Automated Design of Data Visualizations,» de Scientific Visualization - Advances and Challenges, 1994, pp. 88-102.
- [27] R. J. Hendley, N. S. Drew, A. M. Wood y R. Beale, «Narcissus: Visualizing Information,» Int'l Symp. Information Visualization (InfoViz '95), pp. 90-96, 1995.
- [28] E. Tufte, «The Visual Display of Quantitative information,» de Graphics Press, 1983.
- [29] M. Mramor, G. Leban, J. Demsar y B. Zupan, «Visualization-based cancer microarray data classification analysis,» de Bioinformatics, Oxford, 2007, pp. 2147-2154.
- [30] J. A. Ochoa y J. F. Trinidad, «Reconocimiento de patrones,» 2004.
- [31] J. Ruiz-Shulcloper, A. Guzman Arenas y J. F. Martinez-Trinidad, «Selección de Variables y Clasificación Supervisada,» de Enfoque Lógico Combinatorio al Reconocimiento de Patrones I, 1999.
- [32] J. A. Lee, E. Renard, G. Bernard, P. Dupont y M. Verleysen, «Type 1 and 2 mixtures of kullback-leibler divergences as cost functions in dimensionality reduction based on similarity preservation,» Neurocomputing, 2013.
- [33] J. Ham, D. D. Lee, S. Mika y B. Scholkopf, «A kernel view of the dimensionality reduction of manifolds,» Proceedings of the twenty-first international conference on Machine Learning, ACM, p. 47, 2004.
- [34] J. Cook, I. Sutskever, A. Mnih y G. E. Hinton, «Visualizing similarity data with a mixture of maps,» International Conference on Artificial Intelligence and Statistics, pp. 67-74, 2007.
- [35] D. Peluffo-Ordóñez, J. Lee y M. Verleysen, «Generalized kernel framework for unsupervised spectral methods of dimensionality reduction,» IEEE Symposium Series on Computational Intelligence, 2014.
- [36] M. Belkin y P. Niyogi, «Laplacian eigenmaps for dimensionality reduction and data representation,» de Neural computation, vol. 15, 2003, pp. 1373-1396.
- [37] S. T. Roweis y L. K. Saul, «Nonlinear dimensionality reduction by locally linear embedding,» de Science, vol. 290, 2000, pp. 2323-2326.
- [38] G. E. Hinton y S. T. Roweis, «Stochastic neighbor embedding,» de Advances in neural information processing systems, 2002, pp. 833-840.
- [39] J. Harmann, M. P. Murphy, C. S. Peters y P. C. Staecker, «Homotopy equivalence in graph-like digital topological spaces,» 2014.
- [40] D. H. Peluffo-Ordóñez, J. C. Alvarado-Perez, J. A. Lee y M. Verleysen, «Geometrical Homotopy for data visualization,» Computational Intelligence and Machine Learning, 2015.
- [41] M. Lichman, «UCI Machine learning repository,» 2013. [En línea]. Available: <http://archive.ics.uci.edu/ml..>
- [42] S. A. Nene, S. K. Nayar y H. Murase, «Columbia object image library (coil-20),» 1996. [En línea]. Available: <http://www.cs.Columbia.edu/CAVE/coil-20.html>.
- [43] Y. LeCun, L. Bottou, Y. Bengio y P. Haffner, «Gradient-based learning applied to document recognition,» Proceedings of the IEEE 86(11), 1998.
- [44] J. A. Lee, D. H. Peluffo-Ordóñez, and M. Verleysen, «Multi-scale similarities in stochastic neighbour embedding: Reducing dimensionality while preserving both local and global structure,» Neurocomputing, 2015.
- [45] G. Cruz, I. Alberto, C. Cavero, L. Alberto, A. D. Luca, y M. Alfonso, «Implementación de un sistema basado en minería de datos para la obtención de las preferencias estudiantiles de nivel superior para la planificación de materias,» jun. 2011.
- [46] A. Pallasco y E. Mauricio, «Data mining y análisis de datos del proceso de admisión a la educación

- superior en Ecuador”, 2014.
- [47] G. Carrillo y K. Maribel, “Estudio Comparativo de Algoritmos de Predicción Para la Minería de Datos Aplicado al Área Académica FIE-ESPOCH”, jun. 2015.
- [48] N. Barragán, R. Amparo, C. Suica, y M. Yolanda, “Modelo basado en las técnicas de minería de datos aplicada a la detección de ataques en las redes de datos de la Facultad de Informática y Electrónica”, nov. 2014.
- [49] “Aplicación de Técnicas de minería de datos para predecir la deserción de los estudiantes de primer ciclo de la MAD-UTPL - Ordonez Briceño Karla- Informatica.pdf”. [En línea]. Disponible en: <http://dspace.utpl.edu.ec/bitstream/123456789/7897/1/Ordonez%20Brice%C3%B1o%20Karla-%20Informatica.pdf>. [Consultado: 21-mar-2016].
- [50] M. Figueroa, “Minería de datos aplicada a Credit Scoring”, sep. 2006.
- [51] “Computerworld Ecuador - BIG DATA”, Issuu. [En línea]. Disponible en: <https://issuu.com/ekosnegocios/docs/cw278-web>. [Consultado: 21-mar-2016].
- [52] Institución, “Cómo la Big Data impacta en la educación”, Institución. .
- [53] “IBM – Big Data - Colombia”. [En línea]. Disponible en: <http://www-01.ibm.com/software/co/data/bigdata/>. [Consultado: 21-mar-2016].
- [54] “El big data le hace un zoom al consumidor | Revista Líderes”. [En línea]. Disponible en: <http://www.revistalideres.ec/lideres/big-data-le-zoom-consumidor.html>. [Consultado: 21-mar-2016].
- [55] J. A. G. Alonso, J. F. T. Yáñez, J. F. P. Linzán, y Y. P. González, “Seguimiento de la campaña ‘all you need is Ecuador’ en Twitter”, Rev. Publicando, vol. 2, núm. 4, pp. 2–10, ago. 2015.
- [56] “Business Culture - 1310.pdf”. [En línea]. Disponible en: <http://www.ekosnegocios.com/revista/pdfTemas/1310.pdf>. [Consultado: 21-mar-2016].
- [57] Gobierno Nacional del Ecuador, Secretaría Nacional de Planificación y Desarrollo, SENPLADES, Plan Nacional del Buen Vivir 2013 – 2017 <http://www.buenvivir.gob.ec/>

3. FIRMAS DE RESPONSABILIDAD Y SELLO DE LA INSTITUCION



Director del Proyecto

Representante Legal de la
Institución Beneficiaria

4. ANEXOS

ANEXOS

Formato: IFC -001, aprobado por el Comité de Proyectos I+D+i, el 13 de junio del 2011