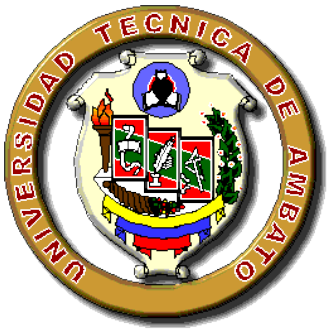
UNIVERSIDAD TÉCNICA DE AMBATO



FACULTAD DE INGENIERÍA EN SISTEMAS ELECTRÓNICA E INDUSTRIAL

MAESTRÍA EN GESTIÓN DE BASES DE DATOS III VERSIÓN

TEMA:

EL USO DE BIG DATA Y SU INCIDENCIA EN LA CALIDAD DE LOS SERVICIOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DE AMBATO.

Trabajo de Investigación, previo a la obtención del Grado Académico de Magíster en Gestión de Bases de Datos

Autor: Ing. Robert Vinicio Vaca Albán

Director: Ing. Edwin Hernando Buenaño Valencia

Ambato – Ecuador

2017

A la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial.

El Tribunal receptor del Trabajo de Investigación presidido por la Ing. Pilar Urrutia, Mg., e integrado por los señores Ing. Jaime Ruíz, Mg., Ing. Edison Álvarez, Mg., Ing. Carlos Núñez, Mg., designados por la Unidad Académica de Titulación de la Facultad de Ingeniería en Sistemas, Electrónica e Industrial de la Universidad Técnica de Ambato, para receptar el trabajo de Investigación con el tema: "EL USO DE BIG DATA Y SU INCIDENCIA EN LA CALIDAD DE LOS SERVICIOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DE AMBATO.", elaborado y presentado por el señor Ing. Robert Vinicio Vaca Albán, para optar por el Grado Académico de Magister en Gestión de Bases de Datos; una vez escuchada la defensa oral del Trabajo de Investigación el Tribunal aprueba y remite el trabajo para uso y custodia en las bibliotecas de la UTA.

Ing. Pilar Urrutia, Mg.
Presidente del Tribunal

Ing. Jaime Ruiz, Mg.
Miembro del Tribunal

Ing. Edison Álvarez, Mg.
Miembro del Tribunal

Ing. Carlos Núñez, Mg.
Miembro del Tribunal

AUTORÍA DEL TRABAJO DE INVESTIGACION

La responsabilidad de las opiniones, comentarios y críticas emitidas en el Trabajo de Investigación presentado con el tema: **EL USO DE BIG DATA Y SU INCIDENCIA EN LA CALIDAD DE LOS SERVICIOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DE AMBATO**, le corresponde exclusivamente a: Ing. Robert Vinicio Vaca Albán, Autor bajo la Dirección de Ing. Edwin Hernando Buenaño Valencia, Mg., Director del Trabajo de Investigación; y el patrimonio intelectual a la Universidad Técnica de Ambato.

Ing. Robert Vinicio Vaca Albán.

CC. 1803006004

AUTOR

Ing. Edwin Hernando Buenaño Valencia

CC. 1802662955

DIRECTOR

DERECHOS DEL AUTOR

Autorizo a la Universidad Técnica de Ambato, para que el Trabajo de Investigación, sirva como un documento disponible para su lectura, consulta y procesos de investigación, según las normas de la Institución.

Cedo los Derechos de mi trabajo, con fines de difusión pública, además apruebo la reproducción de este, dentro de las regulaciones de la Universidad.

Ing. Robert Vinicio Vaca Albán

CC. 1803006004

ÍNDICE GENERAL DE CONTENIDOS

INDICE DE FIGURAS

INDICE DE TABLAS

INDICE DE CUADROS

INDICE DE GRÁFICOS

AGRADECIMIENTO

En primer lugar a Dios que me ha permitido concluir esta etapa muy importante de mi vida.

Un profundo reconocimiento al Ing. Hernando Buenaño por su apoyo durante el desarrollo de éste proyecto de investigación.

A mi esposa e hijos que siempre me dedicaron su apoyo.

A mis padres por sus sabios consejos

A todos quienes me brindaron su conocimiento para la consecución de éste objetivo

DEDICATORIA

Quiero dedicar éste trabajo a mi esposa y mis tres hijos Mishell, Ariel y Mauro, quienes fueron un pilar muy importante para alcanzar esta meta.

UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE INGENIERIA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL / DIRECCIÓN DE POSGRADO

MAESTRIA EN GESTION DE BASES DE DATOS

TEMA:

"EL USO DE BIG DATA Y SU INCIDENCIA EN LA CALIDAD DE LOS SERVICIOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DE AMBATO."

AUTOR: Ing. Robert Vinicio Vaca Albán
DIRECTOR: Ing. Edwin Hernando Buenaño Valencia, Mg.
Fecha: 24 de marzo 2017

RESUMEN EJECUTIVO

El Trabajo de investigación "**El uso de Big Data y su Incidencia en la Calidad de los Servicios Académicos de la Universidad Técnica de Ambato.**", tiene la finalidad de demostrar que toda la información contenida en el Big Data no es clasificable de forma manual, debido a la gran cantidad de información de diferente tipo por parte de la comunidad estudiantil, pasando desde miles de opiniones válidas, serias y constructivas, hasta miles de opiniones subjetivas, irónicas, sarcásticas, mal intencionadas disfrazadas de comentarios. A esto se suma el hecho de que las opiniones son vertidas en redes sociales, las cuales son dinámicas y cada día se tienen miles de opiniones diferentes, convirtiéndose en intratables por el inmenso volumen y dinamismo en el cambio de la información.

Analizar las opiniones vertidas en redes sociales (Big Data), referente a distintos servicios de carácter académico (matriculas, internet, idiomas, etc) que brinda la

Universidad Técnica de Ambato a la comunidad estudiantil, y poder identificar las opiniones reales y descartar aquellas que son subjetivas, irónicas o sarcásticas.

Todo esto con la generación de procedimientos que permitan clasificar los comentarios en mejorables o no, desechando previamente todas aquellas opiniones que contengan emoticones en lugar de palabras, con la finalidad de realizar minería de texto aplicando métodos de agrupamiento como Kmeans y análisis de sentimientos con la ayuda de programas como R y Weka.

Descriptores: Big Data, Minería de Texto, Opiniones, Redes Sociales, Json, Análisis de Sentimientos, Polaridad, Coeficiente Kappa, Matríz de Confusión, Agrupamiento

UNIVERSIDAD TÉCNICA DE AMBATO

FACULTAD DE INGENIERÍA EN SISTEMAS, ELECTRÓNICA E INDUSTRIAL / DIRECCIÓN DE POSGRADO

MAESTRÍA EN GESTIÓN DE BASES DE DATOS

"USE OF BIG DATA AND ITS IMPACT ON THE QUALITY OF ACADEMIC SERVICES OF TECHNICAL UNIVERSITY OF AMBATO."

AUTHOR: Ing. Robert Vinicio Vaca Albán

DIRECTED BY: Ing. Edwin Hernando Buenaño Valencia, Mg.

Date: 24 de marzo 2017

EXECUTIVE SUMMARY

The research paper "the use of Big Data and their impact on the quality of the academics of the Universidad Técnica de Ambato services.", is intended to demonstrate that the information contained in the Big Data is not classifiable manually, due to the large amount of information of different kinds by the student community , from thousands of valid, serious and constructive opinions, even thousands of ill-intentioned sarcastic, ironic, subjective opinions disguised as reviews. A this adds the fact that opinions are expressed in social networks, which are dynamic and every day have thousands of different opinions, becoming untreatable by the immense volume and momentum in the exchange of information.

Analyze the opinions expressed on social networks (Big Data), about various academic services (registration, internet, languages, etc) that provides the Technical University of Ambato to the student community, and to identify the real opinions and discard those that are subjective, ironic or sarcastic.

All this with the generation of procedures that allow to classify the comments in improvements or not, previously discarding all those opinions that contain emoticons instead of words, with the purpose of performing text mining by

applying grouping methods such as Kmeans and analysis of feelings with the Help from programs like R and Weka.

Keywords: Big Data, Text Mining, Opinions, Social Networks, Json, Feelings Analysis, Polarity, Kappa Coefficient, Confusion Matrix, Grouping

INTRODUCCIÓN

En la actualidad las redes sociales se han convertido en una fuente de información importante para las empresas, que permiten el mejoramiento continuo en base a críticas constructivas, de igual manera en el ámbito de la educación día a día estamos rodeados de todo tipo de opiniones favorables y desfavorables referentes a servicios que brinda la Universidad a la comunidad estudiantil.

En la Universidad Técnica de Ambato, se cuenta con la red social Facebook desde la cual se realiza múltiples publicaciones lo que genera respuestas en un volumen alto de información, la cual no puede ser tratado de manera manual por la entropía de cada una de las respuestas, las mismas que muchas de las veces no aportan para un análisis real de los servicios académicos de la UTA.

Este enorme volumen de información para ser clasificado necesitaría de un gran número de personas especializadas en clasificación de información, lo que genera tiempo y costo alto, y por el volumen de información no se realizaría de una manera confiable, lo que conllevaría a tomar decisiones erróneas sobre los servicios académicos que brinda la UTA

La presente investigación se encuentra estructurada de la siguiente manera:

El CAPITULO I.-EL PROBLEMA DE INVESTIGACION, que contiene: El tema de la investigación, Planteamiento del Problema, la Contextualización, Análisis crítico, Prognosis, Formulación del problema, Interrogantes de la investigación, Delimitación del objeto de investigación, Justificación, Objetivos general y Objetivos Específicos.

El Capítulo II.- MARCO TEÓRICO, contiene Antecedentes investigativos, Fundamentación filosófica, Fundamentación legal, Categorías fundamentales, Hipótesis, y Señalamiento de variables de la Hipótesis.

El Capítulo III.-METODOLOGÍA, está formado con Modalidades de investigación, Niveles o tipos de investigación, Población y muestra, Operacionalización de variables y Plan de recolección de la información.

El Capítulo IV.-ANALISIS E INTERPRETACIÓN DE RESULTADOS: contiene el análisis e interpretación de resultados de la encuesta aplicada a los usuarios de los servicios de TI de la Universidad Técnica de Ambato, además se encuentra la Verificación de la Hipótesis.

El capítulo V.- CONCLUSIONES Y RECOMENDACIONES, presenta las conclusiones y recomendaciones de la investigación del problema planteado.

El capítulo VI.- PROPUESTA, contiene el resultado del desarrollo de la solución.

CAPÍTULO I

EL PROBLEMA

1.1 Tema

"EL USO DEL BIG DATA Y SU INCIDENCIA EN LA CALIDAD DE LOS SERVICIOS ACADÉMICOS DE LA UNIVERSIDAD TÉCNICA DE AMBATO".

1.2 Planteamiento del problema

1.2.1 Contextualización

La Universidad Técnica de Ambato durante su vida institucional ha brindado servicios académicos e informáticos a la comunidad universitaria, tales como Internet, matrículas, academia, información, correo electrónico, entre otros como se muestra en la Figura 1.

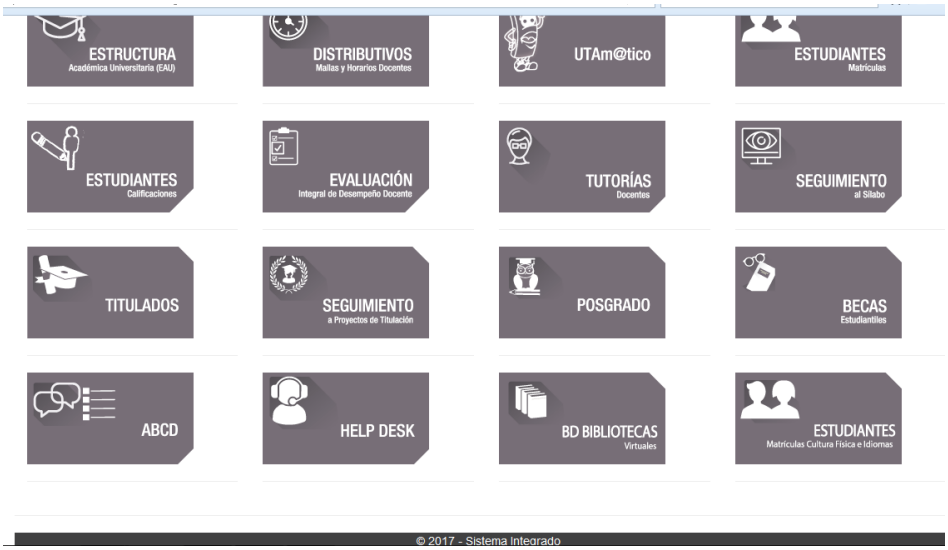


Figura 1. Servicios académicos
Fuente: http://www.uta.edu.ec/v3.0/index.php/es/

Además la UTA posee un sitio en Facebook que le permite obtener opiniones de sus servicios, ver Figura 2:

Figura 2. Sitio Facebook
Fuente:https://www.facebook.com/UniversidadTecnicadeAmbatoOficial

Las opiniones que se escriben en Facebook, son las que conforman el Big Data de este trabajo, puesto que contienen opiniones sobre los servicios académicos de la UTA.

Las encuestas pueden generar miles de opiniones de todo tipo que pueden ser constructivas, sarcásticas, subjetivas, con errores de ortografía, abreviaturas poco entendibles, en donde para clasificarlas se requeriría de un alto número personal especializado en análisis de opiniones, lo cual generaría costos y además por el alto volumen de información existiría un margen de error demasiado grande que derivaría en tomar decisiones equivocadas con respecto a la calidad de los servicios académicos.

El árbol de problemas que el contexto genera se puede observar en la figura3:

1.2.2 Análisis Crítico

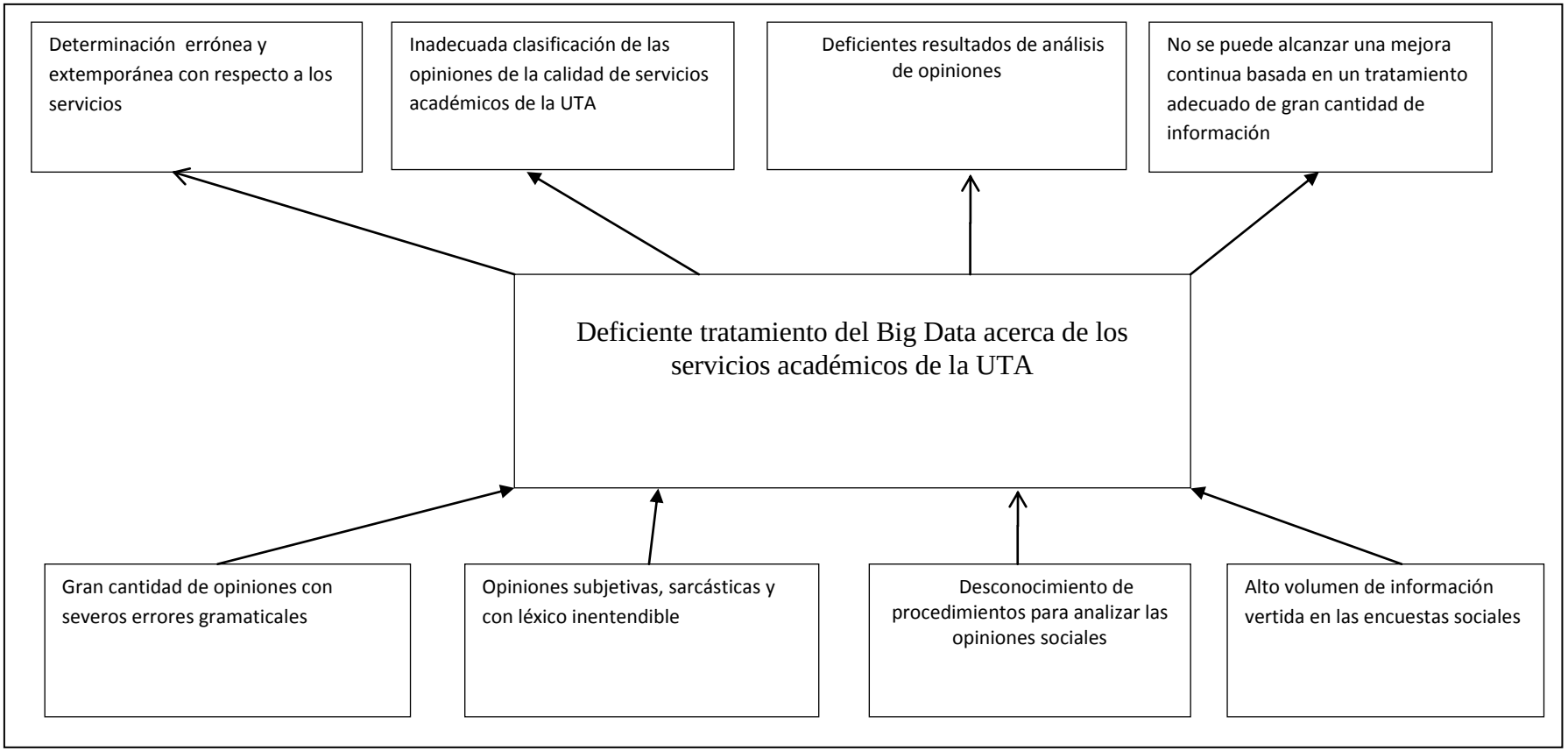


Figura 3: Árbol de Problemas
Elaborado por: Investigador

El Deficiente tratamiento del Big Data acerca de los servicios académicos de la UTA, permite identificar el problema a resolver. Mediante el tratamiento de las opiniones de los estudiantes se puede obtener una definición clara de la calidad de servicios que oferta la universidad, sin embargo, estas opiniones deben ser tratadas y filtradas:

- Gran cantidad de opiniones con severos errores gramaticales, esto implica que deben ser seleccionadas las opiniones válidas y estas ser sometidas al análisis;
- Opiniones subjetivas, sarcásticas y con léxico inentendible, estas opiniones tienen gran contenido en orientación semántica, su carga emocional permite distinguir si es válida o no para el proceso;
- Desconocimiento de procedimientos para analizar las opiniones sociales, no existe un solo proceso para la identificación de opiniones válidas, por lo tanto se debe proponer una metodología acorde al mismo, siendo esta desarrollada en esta tesis;
- Alto volumen de información vertida en las encuestas sociales, esto implica que las opiniones son libres y se manejan bajo información no estructurada (principio propio del Big Data). Se propone estructurar la información, para identificar su carga semántica de opinión y su carga emocional, lo que permite identificar la calidad del servicio.

Aplicados los puntos anteriores, se puede eliminar los errores cometidos en el tratamiento de opiniones sobre la calidad de servicios de la UTA, en un entorno BIG DATA:

- Determinación errónea y extemporánea con respecto a los servicios: Se analiza la información sin un conocimiento previo de los usuarios o de lo que opinan los mismos;
- Inadecuada clasificación de las opiniones de la calidad de servicios académicos de la UTA: si no se realiza un tratamiento adecuado de la información, se tiene un tratamiento inadecuado de la información y probablemente un tratamiento equivoco de los mismos.

- Deficientes resultados de análisis de opiniones: el resultado es analizar la carga semántica y emocional. Si no se tiene un tratamiento metodológico resulta subjetivo el análisis de la opinión. Esto puede ser resuelto consultando a expertos o proponiendo mecanismos de análisis como sucede en este trabajo.
- No se puede alcanzar una mejora continua basada en un tratamiento adecuado del Big Data. Las clasificaciones de las opiniones son dinámicas, y esa es la metodología que se propone. El proceso puede ser aplicado de forma continua con el fin de verificar la validez o mejoras que se tenga que realizar a los servicios ofertados en la UTA.

1.2.3 Prognosis

En caso de no tomar medidas ante el problema, puede ocurrir que la información de las opiniones de los estudiantes en el Big Data acerca de los servicios académicos que presta la UTA se convierta solo en un repositorio inútil de información estática, que nunca será aprovechada correctamente y que derivaría en un desconocimiento total acerca de la calidad de servicios académicos, lo que disminuiría la calidad educativa, administrativa y empresarial de la Universidad, teniendo como principales perjudicados a toda la comunidad universitaria, pues los servicios universitarios dejarían de cumplir su función correcta en la academia.

1.2.4 Formulación del Problema

¿Existe incidencia entre el uso del Big Data y la calidad de los servicios académicos de la Universidad Técnica de Ambato?

1.2.5 Interrogantes de la Investigación

- ✓ ¿Se pueden establecer las características de las opiniones en el Big Data de la UTA?
- ✓ ¿Se puede determinar el nivel de calidad actual de los servicios académicos más utilizados de la UTA, basándose en las opiniones del Big Data?

- ✓ ¿Es posible utilizar el Big Data como un recurso válido que aporte a la toma de decisiones para el mejoramiento continuo de los servicios académicos de la UTA?

1.2.6 Delimitación del Objeto de Investigación

Delimitación de Contenido

Campo: Sistemas Informáticos

Área: Gestión de Base de Datos

Aspecto: Big Data

Delimitación Espacial

La investigación se realizó en la Universidad Técnica de Ambato

Delimitación Temporal

Los datos para la presente investigación se tomaron en el periodo académico septiembre 2016 marzo 2017 y las dos primeras semanas del período marzo septiembre 2017

Unidad de Observación

Opiniones emitidas en el Facebook UTA, sobre servicios académicos.

1.3 Justificación

La presente investigación se justifica porque la UTA está desaprovechando los comentarios y opiniones emitidas directamente por los usuarios primarios de sus servicios académicos, esto se da debido a que no existe una herramienta adecuada para el tratamiento de las opiniones vertidas en el Big Data, derivando que la valiosa información contenida en el Big Data no sea explotada de forma adecuada por lo difícil de su clasificación y manipulación del gran volumen de información existente.

El tratamiento de esta información debería realizarse por parte de personal especializado en análisis de opiniones lo que derivaría en que la UTA, acuda recurrentemente a éste

tipo de talento humano generando altos costos para la Institución, es técnicamente factible por cuanto se tiene la información proporcionada por la encuesta realizada en Facebook. Es factible operativamente porque el investigador tiene los conocimientos necesarios y suficientes para llevar a cabo la misma, y es económicamente factible porque los gastos que genere la investigación correrán a cargo del investigador.

El poder filtrar información y trabajar con las opiniones válidas es **importante** para los directivos. De allí que este trabajo, pretende brindar información válida, a los agentes que toman las decisiones respecto a los servicios virtuales que brinda la universidad. De allí, que es **factible** ya que se cuenta con la información en la red social, que si bien es no estructurada, se propone una metodología para estructurar la información en el Big Data. Y es que es eso, el Big Data la gran cantidad de información y/o opiniones lo que disminuye la capacidad de análisis de la red social, lo cual hace que en este trabajo investigativo, se proponga el mecanismo adecuado y **operativo** para seleccionar, estructurar y manejar las opiniones aun con criterio subjetivo. Todo esto para brindar una cercanía de lo que piensan los usuarios sobre el servicio académico seleccionado.

1.4 Objetivos

1.4.1 Objetivo general

- Establecer la incidencia del Big Data en la determinación de la calidad de los servicios académicos de la Universidad Técnica de Ambato.

1.4.2 Objetivos específicos

- Establecer los mecanismos para tratar la información en el Big Data de la UTA.
- Determinar el nivel de aporte de las opiniones del Big Data en la clasificación de la calidad de los servicios académicos más utilizados de la UTA.
- Proponer una solución para utilizar el Big Data como un recurso válido que aporte a la clasificación de la calidad de servicios académicos de la UTA.

CAPITULO II

MARCO TEORICO

2.1 Antecedentes Investigativos

Algunos trabajos relacionados sobre Análisis de Sentimientos donde se detallan metodologías en diferentes áreas son:

Minería de Opiniones basada en características guiada por Ontologías (Peñalver Martínez, Valencia García, & García Sánchez, 2011)**:** Esta metodología emplea técnicas tradicionales de procesamiento de lenguaje natural junto con procesos de análisis sentimental y tecnologías de la Web Semántica. Los principales objetivos propuestos aquí son mejorar la minería de opinión basada en características empleando ontologías para la selección de las mismas, así como proporcionar un nuevo método para el análisis sentimental basado en el análisis vectorial. Se compone de cuatro módulos principales: el módulo de Procesamiento del lenguaje natural (PLN), el módulo de identificación de características basado en ontologías, el módulo de identificación de la polaridad y el módulo de minería de opiniones.

A Sentimental Education. Sentiment Analysis Using Subjectivity Summarization Based on Minimum Cuts(Pang & Lee)**:** En este trabajo se propone un novedoso método de aprendizaje máquina que aplica técnicas de categorización de texto en las partes subjetivas del mismo. La extracción de estas partes puede ser puesta en práctica usando técnicas eficientes para determinar reducciones (cortes) mínimas en los grafos; facilitando enormemente la incorporación limitada de oración contextual. Por lo tanto, se propone una metodología, que permite emplear primero un detector de subjetividad que determina si cada frase es subjetiva o no; descartando los objetivos, crea un extracto del contenido subjetivo crítico para la clasificación de polaridad predeterminada.

Improving e-learningwithSentimentAnalysis of usersopinions(Kechaou, Ben Ammar, & M Alimi, 2011)**:** Dentro de las tareas fundamentales de la gestión de E-learning es la inspección sistemática de las actitudes y opiniones, con el objetivo de encontrar las necesidades y requisitos tan eficazmente como sea posible. Este enfoque permite a las personas encargadas de E-learning identificar cualquier problema posible

que podría ser encontrado durante su funcionamiento a tiempo y dirigirse a aquellas cuestiones puntuales. De acuerdo a las motivaciones descritas en este trabajo se ha diseñado un algoritmo de clasificación basado en el sentimiento de aprendizaje para dividir las opiniones de un estudiante sobre el servicio del sistema de e-learning en positivo y negativo con el fin de mejorar su rendimiento. En este trabajo, hay tres métodos de selección de la función tradicional **(MI[1], IG[2] y CHI[3])** que han sido investigados y avanzada junto con los **HMM[4]** adecuados y el método de aprendizaje híbrido basado en SVM. Se han realizado experimentos en un corpus de E-learning con un tamaño de 2000 documentos.

OpinionObserver: Analyzing and ComparingOpinionsonthe Web (Bing, Minqing, & Junsheng, 2010)**:** En este trabajo se centra en las opiniones de los clientes en línea de productos. Hace dos contribuciones. En primer lugar, propone un marco novedoso para analizar y comparar las opiniones de los consumidores de productos de la competencia. También se implementa un sistema de prototipo llamado Opinión Observer. El sistema es tal que con un solo vistazo, el usuario es capaz de ver claramente las fortalezas y debilidades de cada producto en la mente de los consumidores de acuerdo a términos de varias características del producto. Esta comparación es útil para los clientes potenciales y los fabricantes de productos. En segundo lugar, se propone una nueva técnica basada en minería de patrón de lenguaje para extraer características del producto en un determinado tipo de revisiones.

2.2 Fundamentación Filosófica

La presente investigación se enmarca en el paradigma Critico propositivo, es crítico por cuanto se realizará un Análisis Crítico del problema y su buscará proponer una solución factible al problema.

2.3 Fundamentación Legal

1**MI.-** Información Mutua.

2**IG.-** Ganancia de la Información.

3**CHI.-** Estadísticas CHI cuadrado.

4**HMM.-** Combinación de los modelos ocultos de Markov.

En la constitución política del Ecuador, en la Sección octava sobre Ciencia, tecnología, innovación y saberes ancestrales, en el Art. 385 menciona "El sistema nacional de ciencia, tecnología, innovación y saberes ancestrales, en el marco del respeto al ambiente, la naturaleza, la vida, las culturas y la soberanía, tendrá como finalidad:

1. Generar, adaptar y difundir conocimientos científicos y tecnológicos.
2. Recuperar, fortalecer y potenciar los saberes ancestrales.
3. Desarrollar tecnologías e innovaciones que impulsen la producción nacional, eleven la eficiencia y productividad, mejoren la calidad de vida y contribuyan a la realización del buen vivir.".

De igual manera el Art. 386 reza: "El sistema comprenderá programas, políticas, recursos, acciones, e incorporará a instituciones del Estado, universidades y escuelas politécnicas, institutos de investigación públicos y particulares, empresas públicas y privadas, organismos no gubernamentales y personas naturales o jurídicas, en tanto realizan actividades de investigación, desarrollo tecnológico, innovación y aquellas ligadas a los saberes ancestrales. El Estado, a través del organismo competente, coordinará el sistema, establecerá los objetivos y políticas, de conformidad con el Plan Nacional de Desarrollo, con la participación de los actores que lo conforman".

Así mismo el Art. 16 sección tercera referente a Comunicación e Información en su literal 2 dice: "El acceso universal a las tecnologías de información y comunicación.", el Art. 18 literal 2 menciona: "Acceder libremente a la información generada en entidades públicas, o en las privadas que manejen fondos del Estado o realicen funciones públicas".

2.4 Categorías fundamentales

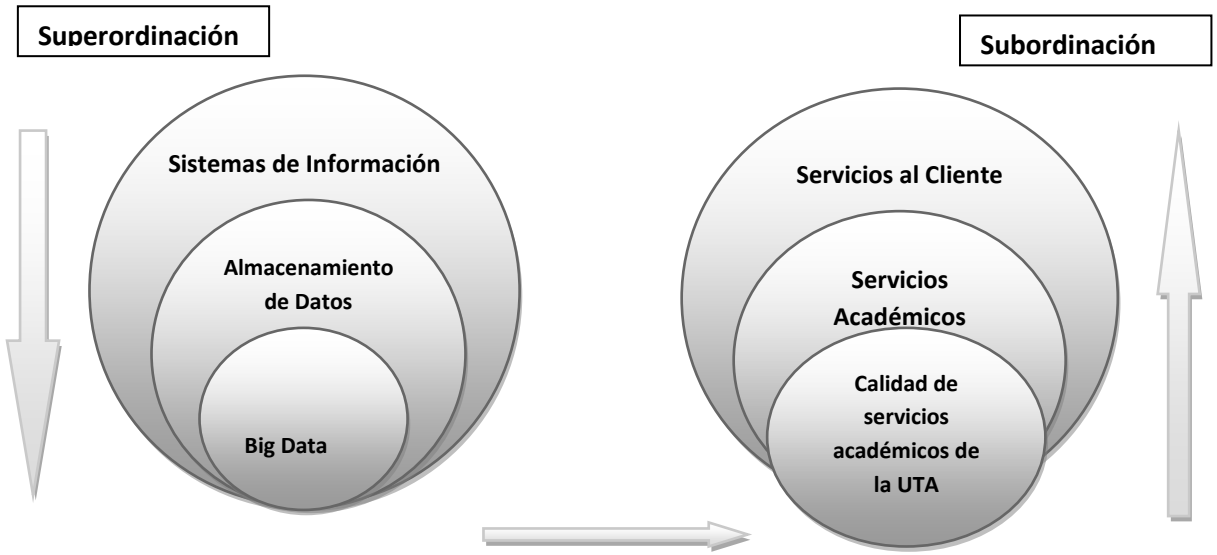


Gráfico 1: Categorías Fundamentales
Elaborado por: Investigador

Constelación de Ideas Variable Independiente

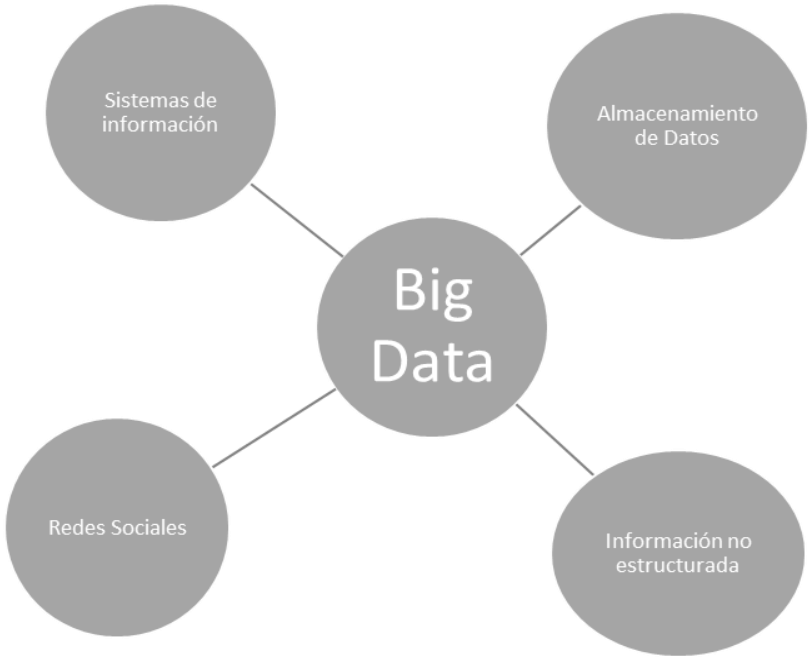


Figura 4: Constelación de Ideas Variable Independiente
Elaborado por: Investigador

Constelación de Ideas Variable Dependiente

Figura 5: Constelación de ideas Variable Dependiente
Elaborado por: Investigador

2.4.1 Categorías Fundamentales de las Variables Independientes

Sistemas de información

Los sistemas de información son mecanismos informáticos que procesan datos de entrada y lo convierten en información valiosa en forma de informes administrativos o estadísticas de rutina para la toma de decisiones a nivel gerencial. Los datos que se toman como insumo para el funcionamiento de los sistemas de información pueden ser formales o informales de tal manera que se puede considerar datos tomados de redes sociales como fuente de información de primera mano para las organizaciones, ya que en ellas es posible descubrir las necesidades y las preocupaciones de los individuos que allí interactúan y transformarlos en soluciones o productos para consumidores potenciales (Kozinets, 2002).

El uso de sistema de información permite a las organizaciones alcanzar de manera más ágil sus objetivos y metas, así como anticiparse en la resolución de problemas que pueden aparecer en el uso de un producto o servicio, y mejorar su competitividad en el ámbito que se desarrolle.

Los sistemas de información son necesarios y esenciales en todo ámbito es así que se puede mencionar algunos ejemplos:

- Un sistema de información en hoteles, permite identificar a los mejores clientes, para obtener la información necesaria para proveer mejores servicios a los mismos.
- En la distribución de cemento, se puede mejorar la entrega de cemento fresco, por medio de dispositivos móviles que analizan el tráfico desde la entrada hasta la entrega del producto.
- La cadena de suministros permite obtener una adecuada comunicación entre proveedores y clientes (Zhang, y otros, 2016)

La tecnología constituye un factor muy importante en la globalización. Algunos países han hecho uso extensivo de la tecnología alcanzando un mejor nivel de vida para sus habitantes.

Actividades en un sistema de información

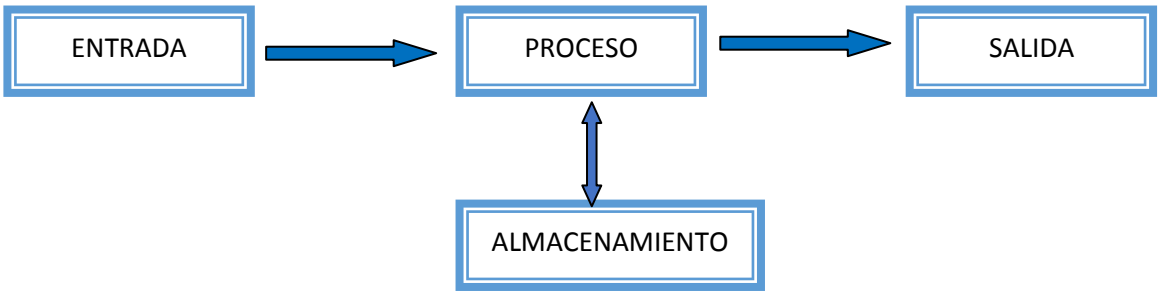


Figura 6: Actividades en un sistema de información
Elaborado por: Investigador

Almacenamiento de Datos

El almacenamiento de datos es una colección de datos orientada a un determinado ámbito, integrado y variable en el tiempo, almacenado en una base de datos o archivos, que ayuda a la toma de decisiones en la entidad en la que se utiliza. Los datos pueden transformar la sociedad, se puede tener grandes conocimientos, por medio de datos extraídos y avanzados, en décadas pasadas los responsables de negocios no contaban con mecanismos o tecnologías para convertir los datos y descubrir la información, hasta la llegada de Big Data. Había distintos orígenes de los datos y los datos del ordenador principal eran inmanejables.

Se obtienen datos estructurados, se toman los datos de distintas fuentes, y se logra centrar a las empresas, el almacenamiento de datos conocido como big data es un organizador de empresas que permite ofrecer información a los clientes, no basta con buscar en páginas web, sino en obtener información de los datos.

Almacenamiento Estructurado

El almacenamiento estructurado se ocasiona en bases de datos en la cual se guardan datos de manera que se tiene mayor independencia, disponibilidad, seguridad, redundancia, eficiencia en la captura, codificación y entrada de datos, garantizando calidad y acceso a los mismos.

En función de la estructura utilizada para construir una base de datos, existen diversos modelos de bases de datos. El modelo de la base de datos define un paradigma de almacenamiento, estableciendo cómo se estructuran los datos y las relaciones entre estos. Las distintas operaciones sobre la base de datos (eliminación o sustitución de datos, lectura de datos, etc.) vienen condicionadas por esta estructura, y existen notables diferencias entre los principales modelos, cada uno de ellos con sus ventajas e inconvenientes particulares. Algunos de los más habituales son los siguientes:

Bases de datos jerárquicas.- Los datos se recogen mediante una estructura basada en nodos interconectados. Cada nodo puede tener un único padre y cero, uno o varios hijos. De este modo, se crea una estructura en forma de árbol invertido en el que todos sus nodos dependen en última instancia de uno denominado *raíz*.

Bases de datos en red.- Este modelo permite la aparición de ciclos en la estructura de la base de datos sin la existencia de un único padre para cada nodo, lo cual permite una mayor eficacia en lo que se refiere a redundancia.

Bases de datos relacionales.- Es el modelo más utilizado, la aparición de éste modelo solucionó los problemas de las bases de datos jerárquicas y de red; el esquema que utiliza es basado en tablas que contienen registros organizados en columnas y filas.

Bases de datos orientados a objetos.- Modelo derivado de los paradigmas de la programación orientada a objetos, extiende las capacidades de datos relacionales, de tal manera que estas pueden contener objetos, facilitando una fácil integración con la arquitectura de los programas para el manejo de base de datos.

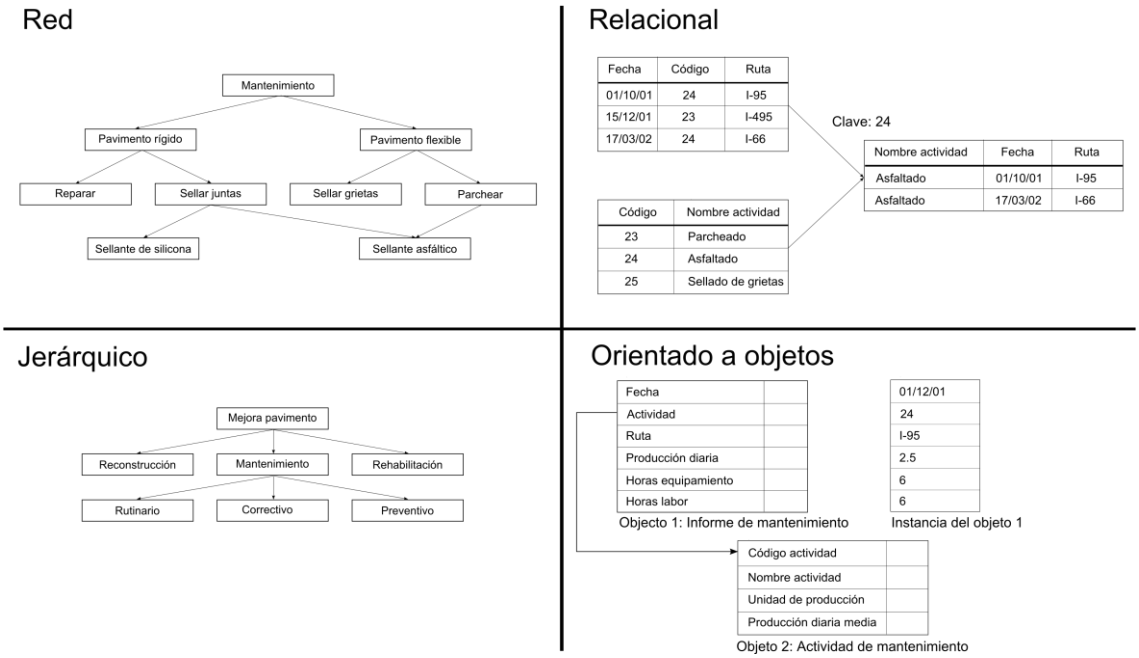


Figura 7: Comparación entre modelos
Fuente:http://volaya.github.io/libro-sig/chapters/Bases_datos.html

Almacenamiento No Estructurado - NoSQL

La expresión "información no-estructurada" se refiere típicamente a aquellos datos que no están organizados bajo el Modelo de Datos Relacional, algunos ejemplos comunes de información no estructurada son los archivos de texto, documentos (PDF, Word), imágenes, audio y video, entre otros

No existe una definición universalmente aceptada de "NoSQL", en general, se refiere a aquellas bases de datos con modelos de consistencia más laxos que los establecidos por las bases de datos relacionales, con el fin de tener una mejor escalabilidad al trabajar con altos volúmenes de información cuando la naturaleza de los datos no requiere un modelo relacional. En un principio, el término se malinterpretó como si el movimiento NoSQL tuviera por objetivo el reemplazar eventualmente a los manejadores de bases de

datos relacionales. A pesar de los beneficios de las bases de datos NoSQL, también tienen desventajas respecto a las bases de datos relacionales, por ejemplo:

- **Poco soporte a JOINs**: Los JOINs en general son caros desde el punto de vista de procesamiento, por eso la mayoría de las bases de datos NoSQL no implementan JOINs, y así logran un mejor desempeño. Para solventar ese problema en una base de datos NoSQL, típicamente se duplican ciertos datos, lo cual, dependiendo del tipo de solución puede ser un inconveniente.
- **Dificultad para manejo de transacciones**: Las bases de datos NoSQL tienen pobre soporte a transacciones, a lo mucho, las soportan a nivel de "registro" en la mayoría de los casos; tampoco soportan transacciones que involucren múltiples entidades. Esto es un problema en el caso de aplicaciones que necesiten mantener la consistencia a lo largo de grupos de registros y múltiples entidades.
- **Nula estandarización**: Cada base de datos NoSQL tiene su propio lenguaje y/o API para acceder y manipular los datos, en muchos casos solo se proporcionan APIs que requieren de programación con el uso de lenguajes imperativos. En contraposición, las bases de datos relacionales utilizan el lenguaje estándar SQL, que es declarativo, ampliamente conocido y de rápido aprendizaje.
- **La flexibilidad del esquema puede ser un problema**: En "malas manos" la flexibilidad en el esquema puede hacer que se lleven a cabo cambios con poco control y a la duplicación de atributos.

En virtud de lo manifestado, en el caso de aplicaciones que utilizan transacciones complejas y tienen un modelo de datos claramente relacional, no es una buena opción una base de datos NoSQL. Un ejemplo típico de una aplicación para la que una base de datos relacional aún es la mejor opción es un "Sistema Bancario Central".

Una definición más apropiada de NoSQL es "NotOnly SQL", es decir, no se trata de descartar el paradigma del modelo relacional, sino de reconocer que existen diferencias en la naturaleza de la información y que, de acuerdo a las características de la misma, así como al uso que se haga de ella, es preferible un paradigma o el otro.

Conforme han evolucionado las soluciones de Big Data se ha buscado emular los beneficios de las bases de datos relacionales al mismo tiempo que se logra escalabilidad a niveles de petabytes en clústers formados por nodos de hardware "commodity". Hive y BigSQL son tecnologías que abordan el problema de implementar (al menos hasta cierto nivel) una base de datos relacional utilizando el paradigma de Big Data.

Base de Datos

Conceptos

Una Base de Datos (BD) es una colección de datos que representa un cierto modelo o abstracción del mundo real (algunas veces llamado el mini - mundo). Los cambios en este mini - mundo son reflejados en la base de datos. La B.D es diseñada, construida y llenada con datos para un propósito específico. Tiene un grupo de usuarios particular, y algunas aplicaciones preestablecidas en las cuáles estos usuarios están interesados. (Universidad Tecnológica de Puebla. (2012).Manual de Asignatura Basado en Competencias Profesionales – Base de Datos I)

Una base de datos es un conjunto de datos persistentes que es utilizado por los sistemas de aplicación de alguna empresa dada.

Componentes de una base de datos

Una base de datos está compuesta principalmente por los siguientes elementos:

Datos: Es la información que llega a la base de datos, constituye la parte esencial que se proporciona al usuario final a través de niveles de abstracción simplificando su interacción; éstos niveles se clasifican en físico, lógico y de vistas, el primero describe cómo se almacenan los datos, el nivel lógico describe qué datos se almacenan y qué relación existe entre ellos, y el nivel de vistas describe sólo parte de la base de datos completa.

Atributos: Constituyen los diferentes campos o características de los objetos que conforman la estructura de la base de datos

Campos: Son unidades mínimas de información a las cuales se puede acceder, cada campo posee un nombre y admite diferentes tipos de datos

Registro: Es un conjunto de campos que contienen datos relacionados al mismo objeto o entidad, posee un número secuencial conocido como número de registro que permite su ubicación, aunque lo recomendable es asignarle un campo clave para facilitar y optimizar su búsqueda.

SGBD

Fray León Osorio Rivera(2008) señala que:

> Un sistema de gestión de bases de datos (SGBD) es un conjunto de programas que permiten el almacenamiento, modificación y extracción de la información en una base de datos, además de proporcionar herramientas para añadir, borrar, modificar y analizar los datos. Los usuarios pueden acceder a la información usando herramientas específicas de interrogación y de generación de informes, o bien mediante aplicaciones al efecto.
>
> Estos sistemas también proporcionan métodos para mantener la integridad de los datos, para administrar el acceso de usuarios a los datos y para recuperar la información si el sistema se corrompe. Permiten presentar la información de la base de datos en variados formatos. La mayoría incluyen un generador de informes. También pueden incluir un módulo gráfico que permita presentar la información con gráficos y tablas.
>
> Hay muchos tipos distintos según cómo manejen los datos y muchos tamaños distintos de acuerdo a si operan en computadoras personales y con poca memoria o grandes sistemas que funcionan en mainframes con sistemas de almacenamiento especiales. Todos los SGBD no presentan la misma funcionalidad, depende de cada producto existen SGBD libres y comerciales entre los más comunes tenemos:
>
> SGBD libres
>
> - PostgreSQL
> - MySQL
> - Firebird
> - SQLite

- Sybase ASE Express Edition para Linux (Edición gratuita para Linux)

SGBD comerciales:

- dBase
- FileMaker
- Fox Pro
- IBM DB2 Universal Database (DB2 UDB)
- IBM Informix
- MAGIC
- Microsoft SQL Server
- Open Access
- Oracle
- Paradox
- PervasiveSQL
- Progress (DBMS)
- Sybase ASE
- Sybase ASA
- Sybase IQ
- WindowBase

Archivos

Un archivo o fichero informático es un conjunto de bits que son almacenados en un dispositivo, el mismo que es identificado por un nombre y la descripción de la carpeta o directorio que lo contiene. A los archivos informáticos se les llama así porque son los equivalentes digitales de los archivos escritos en libros, tarjetas, libretas, papel o microfichas del entorno de oficina tradicional.

Big Data

Los Datos masivos o Big Data es un concepto que hace referencia al almacenamiento de grandes cantidades de datos y a los procedimientos usados para encontrar patrones

repetitivos dentro de esos datos. El fenómeno del Big Data también es llamado datos a gran escala.

Las personas generan gran cantidad de información, siendo esta información la que interesa a las empresas, por lo que han volcado sus esfuerzos a tratar esa información en lo denominado Big Data. Esto tiene diversos objetivos tales como la identificación de patrones, tendencias, prevención de delitos, pronosticar ventas, entre otros. Esta información puede ser recopilada de distintas maneras, para luego ser tratada, con el fin de tomar decisiones sobre lo encontrado en la información. Las empresas han utilizado la big data para ello. WallMart utilizo la big data, esto permitió estar preparada para el huracán catrina, pues al conocer los hechos pudo proveer los productos necesarios para afrontar el fenómeno en la población. Zara la cadena de tiendas españolas, puede predecir gracias al tratamiento con Big Data, la talla de ropa que se venderá en mayoría en una estación del año. Ford, utiliza la Big Data para saber como conducen las personas lo cual ayuda a la compañía en el diseño de nuevos automóviles. En la política utilizan Big Data para proponer campañas políticas, detectando con que público funciona la propaganda en televisión o en radio. Ciudades como Seattle pueden identificar con Big Data el patrón de consumo energético. (Bar-Yam, 2016).

A la hora de clasificar Big Data para que los datos puedan ser interpretados en clave empresarial, hay que tener en cuenta tres factores que podríamos definir como "las tres V":

- **Volumen**: Es el más llamativo por su aumento desmesurado en los últimos años, aunque el menos importante en clave de utilidad para la compañía. Es una consecuencia de las mejoras de las redes de comunicaciones y de las mayores velocidades de los accesos de banda ancha, pero la mayor cantidad de datos por sí sola no aporta un valor añadido. Es la causa que lleva a preocuparse por los otros dos factores.

- **Variedad**: ordenar e interpretar diferentes tipos de datos a la vez puede generar grandes ventajas. Combinar datos de edad, género, estado civil, situación

laboral, situación geográfica, intereses, gustos... permite crear perfiles más precisos de clientes potenciales para realizar campañas de publicidad y marketing segmentadas. Aunque las ventajas de poder ajustar más el punto de mira en el blanco también puede tener una parte negativa, si el cliente percibe una cierta invasión de su intimidad. La sutileza es la gran virtud para que el valor añadido que ofrece la variedad de datos conjuntados no se vuelva en contra.

- **Velocidad**: Se refiere a la vida útil de los datos, no tiene sentido conservar datos cuyo recorrido ha terminado y han quedado obsoletos. Una de las claves para poder almacenar grandes cantidades de datos de forma que sean útiles para la estrategia comercial es que la utilidad de toda la información que se conserva sea vigente. Empresas de según qué sector llegan a descartar hasta el 90% de los datos generados y preservan sólo aquellos que les pueden ofrecer rendimiento.

- **Ventajas competitivas en la gran distribución**: permite actualizar, optimizar y afinar inventarios en función de la demanda en tiempo real. El datamining o extracción de datos posibilita el análisis del comportamiento de los clientes, fijar los precios en consecuencia u ofrecer los incentivos adecuados para atraer a los clientes. Es posible realizarlo mediante la digitalización de elementos como los almacenes o las cajas registradoras.
- **Mejora de la eficiencia y los costes**: el análisis del Big Data puede acelerar la velocidad con que se desarrolla un producto. También permite compartir datos de forma rápida y realizar simulaciones de producto. En algunos sectores, los plazos de desarrollo se han llegado a reducir entre el 30% y el 50%.
- **Mejora de la gestión empresarial**: además de optimizar la cadena de suministro y el inventario, el Big Data puede ser útil para reducir el ciclo de conversión de efectivo, controlar factores de riesgo y tomar decisiones empresariales que pueden virar el futuro de la empresa fundamentadas en datos recogidos en tiempo real.

- **Almacenamiento en la nube**: uno de los problemas para gestionar altos volúmenes de datos es el elevado costo de la infraestructura de almacenamiento. Muchos proveedores de almacenamiento masivo de datos alquilan potentes servidores a los que se puede acceder en línea, y ponen a disposición del cliente como solución almacenarlos en una especie de nube. El resultado es que se puede acceder a ella mediante aplicaciones diseñadas para manejar grandes volúmenes de datos y se pueden obtener soluciones a menudo en tiempo real de forma sencilla.

Big Data (opiniones en redes sociales)

Las redes sociales se han introducido recientemente en la vida de muchas personas que antes eran ajenas al fenómeno de Internet. No es extraño oír hablar por la calle de Facebook y no necesariamente entre los más jóvenes. La extraordinaria capacidad de comunicación y de poner en contacto a las personas que tienen las redes ha provocado que un gran número de personas las esté utilizando con fines muy distintos. Se utilizan para encontrar y entablar diálogo con amistades perdidas tiempo atrás, para debatir sobre los temas más variados, apoyar causas de todo tipo, organizar encuentros de amigos, ex-compañeros de estudios o para dar a conocer congresos y conferencias, a través de los cuales no sólo se dan detalles sobre el encuentro, sino que las personas pueden confirmar su asistencia o ausencia al evento.

El mundo educativo no puede permanecer ajeno ante éstos fenómenos sociales que están cambiando la forma de comunicación entre las personas. El sistema educativo trabaja fundamentalmente con información, carecería de sentido utilizar sistemas de transmisión y publicación de la misma basada en aquellos que se utilizaban a principios y mediados del siglo XX sin incorporar aquello que la sociedad ya está usando como parte de su vida cotidiana. La educación debe formar las personas para aquello que serán y en lo que trabajarán dentro de diez años

Tipos de servicios de redes sociales

Las redes sociales son estructuras compuestas por personas u otras entidades humanas las cuales están conectadas por una o varias relaciones que pueden ser de amistad, laboral, intercambios económicos o cualquier otro interés común

Redes sociales estrictas

Este tipo de redes son las que presentan un mayor valor en su aplicación educativa debido a su inespecificidad, con lo que se pueden adaptar libremente según las necesidades.
Debemos distinguir claramente entre las redes sociales que se descargan de Internet y se ejecutan en los servidores propios del centro educativo y aquellas que están alojadas en servidores de terceros en manos de empresas especializadas.

Microblogging

Las redes sociales basadas en microblogging(también llamadas de Nanoblogging) son las que se basan en mensajes cortos de texto.

El ejemplo más conocido de este tipo de redes es Twitter (http://twitter.com) aunque existen muchas más. El problema más importante que tienen de cara a la educación es la limitación en la creación de objetos digitales, ya que se basan principalmente en el texto y, aunque muchos servicios permiten la inserción de vídeos, imágenes, archivos y otro tipo de elementos, no tienen las posibilidades de las otras redes sociales.

Redes sociales completas

Este tipo de redes, a las que se ha denominado completas para diferenciarlas de las basadas en microblogging, permiten una mayor comunicación e interacción entre sus miembros, además se pueden compartir todo tipo de objetos digitales además del texto.

El paradigma de estas redes se encuentra en Facebook o Tuenti donde los usuarios establecen lazos de amistad mutua lo cual les da acceso al perfil del otro usuario, así

como ponerse en contacto con él de muchas formas distintas (a través de comentarios en su muro, en sus fotos, enviándole regalos, juegos, etc.).

Las redes sociales de este tipo tienen que cumplir dos características básicas para ser aptas y útiles en educación. La primera es la posibilidad de crear redes cerradas para todo el que no esté registrado y la segunda es la posibilidad de crear grupos o subredes dentro de la propia red.

Usos educativos de las redes sociales completas

La plasticidad de las redes hace que sus aplicaciones sean tantas como docentes las utilicen. Existen muchas formas de usar las redes sociales en educación y aquí mostraremos únicamente algunas de las más generales y habituales.

Redes de asignaturas

En algunos casos se crea una red específicamente para una asignatura, con la finalidad de establecer un diálogo a partir de la red, consultar dudas, realizar trabajos, etc. Tal como se ha comentado anteriormente, las redes poco numerosas desaprovechan las capacidades sociales que tienen, así pues, aunque es posible utilizarlas de este modo lo más aconsejable será intentar usar las redes con más asignaturas, alumnos y profesores.

Redes de centros educativos y grupos para crear comunidades internas

Sin duda este es el uso más fructífero para las redes sociales educativas. Un centro educativo, sea un colegio, instituto, academia o universidad, en una única red social crea un sentimiento de pertenencia a una comunidad real. Las diferentes asignaturas, tutorías o agrupaciones de cualquier otro tipo se pueden realizar a través de los grupos internos de la red. Los casos que se describen a continuación pertenecen a este modelo de red.

Grupos como lugar de consulta de dudas y de actividad académica de una asignatura

Una posible forma de usar los grupos de las redes sociales es como un sitio privado para los alumnos de una asignatura y su profesor. Un lugar al que los alumnos pueden acudir para estar en contacto con su profesor, ya sea para preguntar sobre la materia, consultar notas de los exámenes, etc.

Grupo como tablón de anuncios de la asignatura

Se puede utilizar el grupo como lugar donde colocar todas las tareas, trabajos o deberes que deben realizar los alumnos. Los servicios de redes sociales que disponen grupos con blogs son ideales para desempeñar esta función ya que cada día el profesor puede publicar las tareas del día en el blog del grupo.

Grupos de alumnos

Los alumnos forman grupos para colaborar en la realización de trabajos, además utilizan foros de discusión, muros y demás herramientas para organizarse, compartir información entre todos los compañeros y construir el trabajo final de manera conjunta.

2.4.2 Categorías Fundamentales de la Variable dependiente

Calidad de servicios académicos

El establecimiento de estándares de calidad es un factor importante para las organizaciones ya que sirve de punto de referencia en la búsqueda de la permanencia en el mercado de las mismas debido a la competitividad que esto representa. Las instituciones educativas forman parte de empresas de servicio, ya que sus productos o procesos no son tangibles y la calidad de los mismos depende de la percepción de los clientes (estudiantes). Siendo el nivel de servicios prestado una medida de calidad. (Santamaría P. & Mejías A., 2013)

Las instituciones educativas de nivel superior, han encontrado en la calidad un factor importante a considerar para la búsqueda de su permanencia en el tiempo. Esta calidad podría identificarse en los atributos que debe poseer un graduado para poder cumplir

con los requerimientos de los empresas o consumidores; para el caso de una empresa productora de bienes, esta estará siendo evaluada de acuerdo con las características y enfoques que se tengan de los procesos y productos desarrollados

Existen diferentes perspectivas desde las cuales se ve la calidad con el objetivo de evaluar el papel que desempeña la educación superior en las distintas partes de la organización, entre estas se encuentran, perspectiva de que el usuario se base en que la calidad se determina de acuerdo con lo que el cliente quiere, con base en el valor; siendo esta la relación con el precio y la satisfacción, con base en la manufactura, este es el resultado deseable de la práctica de ingeniería y manufactura o la conformidad con las especificaciones, entre otras (Evans y Lindsay, 2008).

Deming (1989), plantea que la dificultad para definir la calidad se encuentra en traducir las futuras necesidades de los consumidores en características medibles permitiendo que el producto sea diseñado y sea el resultado del precio que el cliente esté dispuesto a pagar. Así mismo, Feigenbaum (1986), señala que la calidad es determinada por el cliente, esta se basa en la experiencia real del cliente con el producto o servicio, relacionada con los requisitos declarados o no declarados y conscientes o simplemente detectada, técnicamente operativa o totalmente subjetiva y siempre lo que representa un objetivo móvil en un entorno competitivo.(Santamaría P. & Mejías A., 2013)

Servicios Académicos

Constituyen actividades desarrolladas en el área de la academia con la finalidad de satisfacer a los involucrados en el proceso educativo, pudiendo ser: Matrículas, record académico, internet, bibliotecas virtuales, correo electrónico institucional, aulas virtuales, programas de educación continua, entre otros.

La Universidad Técnica de Ambato adopta las actividades académicas no solo de manera conceptual sino con un verdadero sentido de servicio hacia la comunidad

universitaria (estudiantes y docentes), mediante aplicaciones informáticas que permiten agilitar los procesos vinculados a la educación superior.

Dentro de los servicios académicos informáticos más utilizados se puede mencionar:

Matrículas: Servicio académico informático brindado al sector estudiantil de la Universidad Técnica de Ambato, el mismo que permite realizar la matrícula en línea durante cada período de matrículas.

Record Académico: Servicio web que permite consultar la trayectoria académica del estudiante

Internet: Servicio universitario que permite navegación wifi a estudiantes, docentes y personal administrativo, el mismo que se encuentra distribuido en todos los predios de la UTA.

Correo electrónico institucional: Brinda acceso a cuentas creadas en office 365 facilitando la comunicación institucional.

Bibliotecas virtuales: Permite acceder a bases de datos de objetos digitales facilitando la búsqueda de información bibliográfica.

Servicio al cliente

Es un conjunto de actividades que buscan satisfacer las necesidades del cliente, las mismas que incluyen diversas actividades desempeñadas por un gran número de personas (funcionarios, empleados, directivos). Para la Universidad Técnica de Ambato, los clientes representan los estudiantes y las actividades que buscan satisfacer las necesidades de los mismos están vinculadas al proceso enseñanza – aprendizaje, cuyos actores involucrados en brindar un excelente servicios son: Autoridades, docentes, administrativos y personal de servicios.

Como un ejemplo se servicio al cliente en la educación se puede considerar los cursos en línea MOOCS que implica que los profesores pueden ver cuando los estudiantes regresan a sus lecciones, se han detenido o avanzan, lo que permite aprender si es necesario preparar o mejorar la clase o detenerse en una lección determinada.

Si las universidades pueden monitorear el estudio de un estudiante, se puede proponer que las tutorías sean personalizadas. Clases de matemáticas que se dictan por ordenador, y los profesores no son los únicos que ayudan en los estudios, sino sistemas recogen, datos del estudiante durante el proceso de enseñanza aprendizaje. Un estudiante puede avanzar sin problemas en una asignatura, mientras a otra le falta más, esto implica que los sistemas de atención al cliente se personalicen para cada uno adaptándose a la enseñanza. Los profesores pueden procesar como avanza los alumnos, gracias a los datos que le proporciona el sistema de atención al estudiante (Scull, Thorup, & Howell, 2016).

2.5 Hipótesis

El uso de Big Data permite determinar la calidad de los servicios académicos de la Universidad Técnica de Ambato

2.6 Señalamiento de variables e Hipótesis

Variable Independiente

- Big Data

Variable Dependiente

- Calidad de servicios académicos

CAPÍTULO III

METODOLOGÍA

3.1 Enfoque

La presente investigación es cuali-cuantitativa porque utiliza parámetros de medición de las opiniones en Big Data, lo que permitirá la toma de decisiones para mejorar la calidad de los servicios académicos.

3.2 Modalidad de Investigación.

Investigación Bibliográfica

La modalidad de investigación será bibliográfica porque utilizará fuentes como: libros, documento, revistas, artículos científicos, documentos publicados en internet que aportarán para la construcción del marco teórico de la investigación.

Investigación de Campo

La investigación también tendrá la modalidad de campo porque se buscará obtener la información de la variable independiente, acerca del uso del Big Data y la variable dependiente calidad de los servicios académicos.

3.3 Niveles o tipos de Investigación

Investigación Exploratoria

La presente investigación será Exploratoria porque buscará destacar los aspectos fundamentales de la calidad de servicios en la Universidad Técnica de Ambato y encontrar los procedimientos adecuados para elaborar una investigación posterior con una posible solución, en relación a distinguir opiniones válidas en Big Data.

Investigación Descriptiva

También será descriptiva porque se utilizara métodos de análisis para caracterizar el objeto de estudio.

Investigación Explicativa

Será Explicativa porque utiliza los métodos analítico y sintético, en conjugación con los métodos deductivo y inductivo mediante el establecimiento de relaciones causa-efecto.

Investigación Correlacional

Será también correlacional, porque buscara medir el grado de relación entre la variable independiente el uso del Big Data y la variable dependiente calidad de los servicios académicos.

3.4 Población y Muestra

El número de seguidores que tiene el sitio oficial de Facebook de la Universidad Técnica de Ambato es superior a 25000 de los cuales 15000 son estudiantes de nivelación, pregrado, posgrado y centro de idiomas los cuales constituyen la población de la presente investigación.

Como la población supera las 100 personas, para determinar la muestra se aplica la siguiente fórmula

$$n= \frac{k^2*p*q*N}{(e^2*(N-1))+k^2*p*q}$$

N: es el tamaño de la población o universo.

k: es una constante que depende del nivel de confianza que asignemos. El nivel de confianza indica la probabilidad de que los resultados de nuestra investigación sean ciertos: un 95,5 % de confianza es lo mismo que decir que nos podemos equivocar con una probabilidad del 4,5%.

Los valores k más utilizados y sus niveles de confianza son:

Tabla 1:Valores k

K	1,15	1,28	1,44	1,65	1,96	2	2,58
Nivel de confianza	75%	80%	85%	90%	95%	95,5%	99%

Elaborado por: Investigador

95% (1.96) el valor mínimo aceptado para considerar la investigación como confiable.

e: es el error muestral deseado, que es la diferencia que puede haber entre el resultado que obtenemos preguntando a una muestra de la población y el que obtendríamos si preguntáramos al total de ella.

5% (0.5) es el valor estándar usado en las investigaciones

p: es la proporción de individuos que poseen en la población la característica de estudio. Este dato es generalmente desconocido y se suele suponer que p=q=0.5 que es la opción más segura.

q: es la proporción de individuos que no poseen esa característica, es decir 1-p.

n: es el tamaño de la muestra a obtener.

Entonces aplicando la fórmula se obtiene la muestra a utilizar de 375

N= 15000

K= 1.96

e= 5%

P= 0.5

Q= 0.5

n= 375

3.5 OPERACIONALIZACIÓN DE LAS VARIABLES

3.5.1 Operacionalización de la Variable Independiente:

Variable Independiente: El Big Data

CONCEPTUALIZACIÓN	DIMENSIONES	INDICADORES	ITEMS BÁSICOS	TÉCNICAS E INSTRUMENTOS
El Big data es el conjunto masivo de opiniones disponible en internet generalmente no estructurada	Conjunto masivo de opiniones Disponible en Internet No estructurada	Cantidad de opiniones Tipos de redes sociales. Opiniones libres de estructura	Cree que el big data al almacenar grandes volúmenes de información puede generar información para toma de decisiones gerenciales Cree que las redes sociales son un mecanismo adecuado para la captura masiva de comentarios que no se vean atados a un criterio de quien publica Posee usted una cuenta en la red social Facebook Piensa que el uso masivo de información	Encuesta Cuestionario

			afecta a los sistemas informáticos	

Cuadro 1: Operacionalización de la Variable Independiente
Elaborado por: Investigador

3.5.2 Operacionalización de la Variable Dependiente:

Variable Dependiente: Calidad de servicios académicos

CONCEPTUALIZACIÓN	**DIMENSIONES**	**INDICADORES**	**ITEMS BÁSICOS**	**TÉCNICAS E INSTRUMENTOS**
La calidad es un elemento que permite medir la eficiencia de un valor en este caso de un servicio académico.	Medición de eficiencia de un servicio académico**.**	Opiniones validadas	Cree que los servicios académicos basados en tecnologías de la información de la UTA son suficientes Conoce los servicios académicos basados en tecnología que oferta la UTA	Encuesta Cuestionario

Cuadro 2: Operacionalización de la Variable Dependiente
Elaborado por: Investigador

Técnicas e instrumentos

Las técnicas e instrumentos que se utilizaron para la recolección de la información fueron: Encuesta / Cuestionario

Plan de Recolección de Información

PREGUNTAS BÁSICAS	EXPLICACIÓN
¿Para qué?	Para alcanzar los objetivos de la investigación
¿De qué personas u objetos?	Usuarios de redes sociales (Facebook)
¿Sobre qué aspectos?	Indicadores expuestos en la Matriz de operacionalización de variables
¿Quién, Quiénes?	Robert Vinicio Vaca Albán
¿Cuándo?	Último mes a partir de la aprobación de proyecto
¿Dónde?	Universidad Técnica de Ambato
¿Cuántas veces?	Una vez por mes
¿Qué técnicas de recolección?	Encuesta / Cuestionario
¿Con qué?	Encuesta
¿En qué situación?	En condiciones normales

Cuadro 3: Recolección de Información
Elaborado por: Investigador

3.6 Plan de Recolección de Información

Para la recolección de la información se realizaron las siguientes actividades:

- Determinar el servicio académico a ser analizado
- Diseño de la Encuesta online
- Invitación enviada usando el mail institucional
- Aplicación del instrumento vía web

3.7 Plan de Procesamiento de Información

Los datos recogidos se transforman con los siguientes procedimientos.

- Tabulación o cuadros según variables de cada hipótesis: cuadros de una sola variable, cuadro con cruce de variables, etc.
- Estudio estadístico de datos para presentación de resultados.
- La presentación de datos puede hacerse siguiendo los siguientes procedimientos:

    - Representación escrita
    - Representación semitabular
    - Representación tabular
    - Representación gráfica
- Análisis e interpretación de resultados
- Análisis de los resultados estadísticos, destacando tendencias o relaciones fundamentales de acuerdo con los objetivos e hipótesis.
- Interpretación de los resultados, con apoyo del marco teórico, en el aspecto pertinente.
- Comprobación de hipótesis Para la verificación estadística conviene seguir la asesoría de un especialista.
- Establecimiento de conclusiones y recomendaciones

CAPÍTULO IV

ANALISIS E INTERPRETACION DE RESULTADOS

4.1 Análisis de los Resultados

Los datos obtenidos de la investigación fueron ordenados y procesados, mediante el análisis, para luego ser valorados, mediante la utilización de estadística descriptiva, con cuadros y gráficos, en las cuales constan los respectivos análisis e interpretación tomando en consideración los objetivos de las interrogantes y el marco teórico. A continuación se detalla los resultados obtenidos en la encuesta.

4.1.1 ENCUESTA APLICADA A LOS USUSARIOS QUE UTILIZAN LOS SERVICIOS ACADEMICOS DE LA UTA BASADOS EN TECNOLOGIAS DE LA INFORMACION.

PREGUNTA 1. ¿Cree que el big data al almacenar grandes volúmenes de información puede generar información para toma de decisiones gerenciales?

Tabla 2:Pregunta 1

Alternativas	Frecuencias	Porcentajes
SI	255	68,00%
NO	120	32,00%
TOTAL	375	100,00%

Elaborado por: Investigador

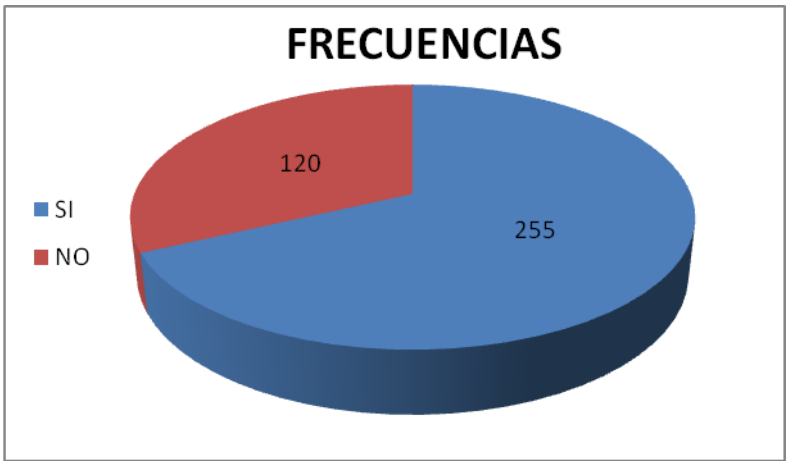


Gráfico 1: Representación Pregunta 1
Elaborado por: Investigador

Análisis

De un total de 375 personas que llenaron la encuesta, un 68% opina que el big data al almacenar grandes volúmenes de información SI puede generar información para toma de decisiones gerenciales, mientas que un 32% indican que NO.

Interpretación

Conforme a los datos obtenidos se puede decir que la mayoría de los usuarios opinan que el big data al almacenar grandes volúmenes de información SI puede generar información para toma de decisiones gerenciales, sin embargo, el número que indican que NO es bastante representativo, lo cual puede deberse al desconocimiento técnico del tema de Big Data.

PREGUNTA 2. ¿Cree que las redes sociales son un mecanismo adecuado para la captura masiva de comentarios que no se vean atados a un criterio de quien publica?

Tabla 3:Pregunta 2

Alternativas	Frecuencias	Porcentajes
SI	227	60,50%
NO	148	39,50%
TOTAL	375	100,00%

Elaborado por: Investigador

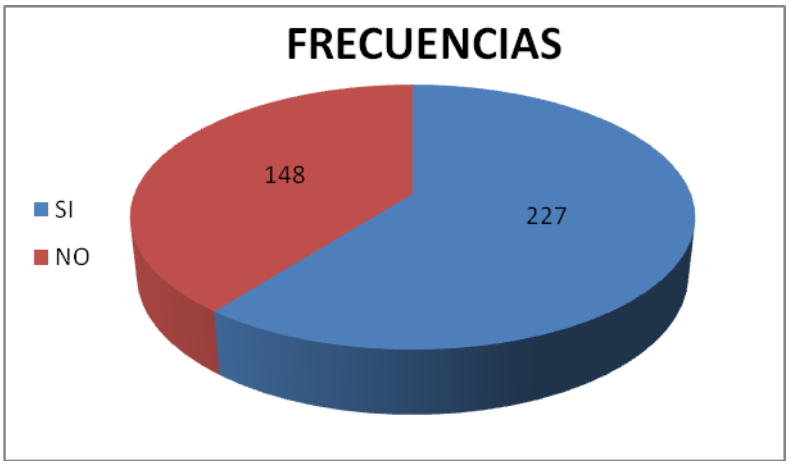


Gráfico 2: Representación Pregunta 2

Elaborado por: Investigador

Análisis

De las 375 personas que se les aplico la encuesta, un 60.5% opina que las redes sociales SI son un mecanismo adecuado para la captura de información a través de los comentarios y opiniones que hacer en la red, mientas que un 39.5% indican que NO.

Interpretación

Según los datos obtenidos se puede interpretar que la mayoría de las personas opinan que las redes sociales son un medio adecuado para levantar información basadas en las opiniones que dejan en un tema propuesto, además que estas opiniones son libres y no se someten a criterios sesgados que suelen esconderse en otros instrumentos de aplicación hacia los usuarios.

PREGUNTA 3. ¿Posee usted una cuenta en la red social Facebook?

Tabla 4: Pregunta 3

Alternativas	Frecuencias	Porcentajes
SI	375	100,00%
NO	0	0,00%
TOTAL	375	100,00%

Elaborado por: Investigador

Gráfico 3: Representación Pregunta 3

Elaborado por: Investigador

Análisis

De las 375 personas que se les aplico la encuesta, todos contestan que poseen una cuenta de Facebook. Es decir el 100% de las personas que se les aplica el instrumento dicen que SI tienen una cuenta activa, frente a un 0% que dice que NO.

Interpretación

A través de los datos obtenidos se puede verificar que la totalidad de las personas a quienes se les aplico la encuesta poseen una cuenta en la red social de Facebook, lo que se puede interpretar como una ventaja para el presente proyecto ya que los usuarios de los servicios TI de la UTA tienen un acceso diario a las publicaciones que se le hace acerca de los servicios en oferta y pueden opinar libremente de los mismos.

PREGUNTA 4. ¿Cree que los servicios académicos basados en tecnologías de la información de la UTA son suficientes?

Tabla 5: Pregunta 4

Alternativas	Frecuencias	Porcentajes
SI	9	2,40%
NO	366	97,60%
TOTAL	375	100,00%

Elaborado por: Investigador

Gráfico 4: Representación Pregunta 4

Elaborado por: Investigador

Análisis

De las 375 personas que se les aplico la encuesta, un 2,4% contestan que los servicios académicos basados en tecnologías de la información de la UTA SI son suficientes. Mientras un 97,6% indican que no son suficientes.

Interpretación

A través de los datos obtenidos se puede indicar la mayoría de las personas dicen que los servicios académicos basados en tecnologías que oferta la UTA NO son suficientes para satisfacer la demanda de los usuarios, lo que hace pensar que se debe continuar mejorando o aumentando los mismos.

PREGUNTA 5. ¿Piensa que el uso masivo de información afecta a los sistemas informáticos?

Tabla 6: Pregunta 5

Alternativas	Frecuencias	Porcentajes
SI	167	44,50%
NO	208	55,50%
TOTAL	375	100,00%

Elaborado por: Investigador

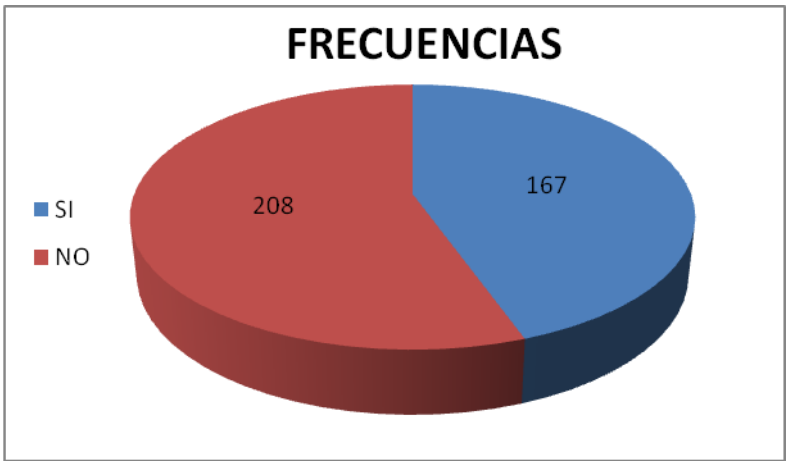


Gráfico 5: Representación Pregunta 5

Elaborado por: Investigador

Análisis

De las 375 personas que se les aplico la encuesta, un 44,5% contestan que el uso masivo de la información SI afecta a los sistemas informáticos de la UTA. Mientras un 55,5% indican que NO afectan.

Interpretación

A través de los datos obtenidos se puede indicar la mayoría de las personas entienden que el uso masivo de información afecta a los sistemas informáticos, de tal forma que a través de esta pregunta nos podemos dar cuenta el conocimiento técnico que poseen los encuestados, con la finalidad de saber el criterio a la hora de haberles aplicado la encuesta.

PREGUNTA 6. ¿Conoce los servicios académicos basados en tecnología que oferta la UTA?

Tabla 7: Pregunta 6

Alternativas	Frecuencias	Porcentajes
SI	306	81,60%
NO	69	18,40%
TOTAL	375	100,00%

Elaborado por: Investigador

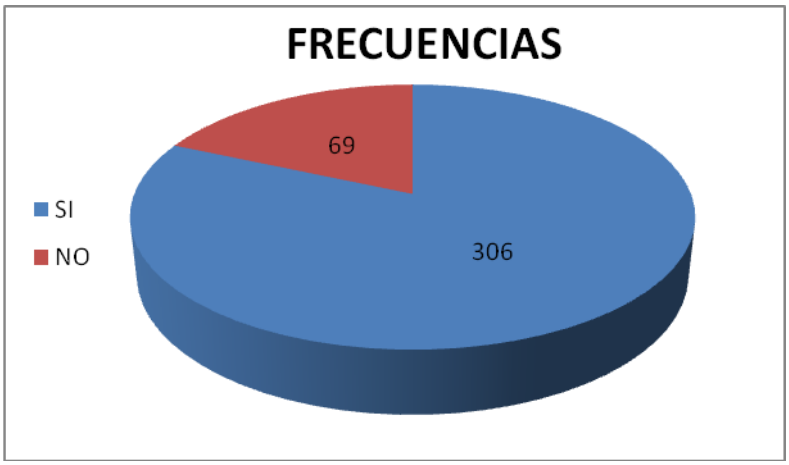


Gráfico 6: Representación Pregunta 6

Elaborado por: Investigador

Análisis

De las 375 personas que se les aplico la encuesta, un 81,6% contestan que SI conocen los servicios académicos ofertados por la UTA. Mientras un 18,4% indican que NO los conocen.

Interpretación

Por medio de las respuestas obtenidas a esta pregunta se puede indicar que la mayoría de personas encuestadas conoce los servicios académicos ofertados por la UTA en base a tecnologías de la información y una importante minoría no conoce de estos servicios, lo que hace pensar que se debe llevar una campaña informativa para que los usuarios conozcan.

4.2 Verificación de la hipótesis

Para la verificación de la hipótesis se utilizará el estadístico del chi-cuadrado el mismo que nos permitirá, obtener la información pertinente para aceptar o rechazar la hipótesis planteada.

Hipótesis.- "El uso de Big Data permite determinar la calidad de los servicios académicos de la Universidad Técnica de Ambato".

4.2.1 Combinación de frecuencias

Para establecer la correspondencia de las variables se eligió dos preguntas de las encuestas, una por cada variable de estudio, lo que permitió efectuar el proceso de combinación.

4.2.2 Planteo de hipótesis HIPÓTESIS NULA H0.

El uso de Big Data **NO** permite determinar la calidad de los servicios académicos de la Universidad Técnica de Ambato

HIPÓTESIS ALTERNA H1. El uso de Big Data SI permite determinar la calidad de los servicios académicos de la Universidad Técnica de Ambato.

4.2.3 Definición del nivel de significación: El nivel de confianza escogido para el presente trabajo es del 95% ($\propto$=0.05)

4.2.4 Definición de la población

Se trabajó con una población calculada de 375 estudiantes que son seguidores del sitio oficial de Facebook de la UTA a quienes se les aplicó un cuestionario vía online usando los medios con que cuenta la Institución (Forms de Office 365).

4.2.5 Especificaciones del estadístico

De acuerdo a la contingencia se utilizará la siguiente fórmula:

$$x^2 = \sum \left[(fo - fe)^2 / fe \right]$$

Dónde:

x^2 : Chi o Jí cuadrado

fo : Frecuencias observadas

fe : Frecuencias esperadas

$\sum$: Sumatoria

4.2.6 Especificación de las regiones de aceptación y rechazo

Para decidir sobre estas regiones, se determinará los grados de libertad, conociendo que el cuadro está formado por dos filas y dos columnas

gl =(f -1)* (c -1)

gl =(6-1)* (2-1)

gl =(5)* (1)

gl =5

Dónde:

c = Columnas

f = Filas

gl = grados de libertad

Por lo que con 5 gl y un nivel de 0,05 tenemos en la tabla de x^2 el valor de 11.07.

La representación sería:

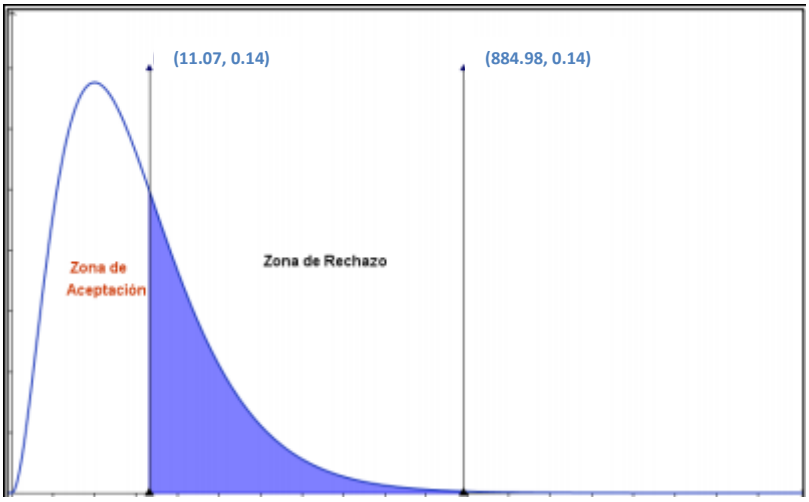


Gráfico 1. Representación estadística calidad de servicios académicos

Elaborado por: Investigador

4.2.6 Recolección de los datos de cálculo de los estadísticos

Tabla 8: Frecuencias Observadas

FRECUENCIAS OBSERVADAS			
PREGUNTAS	**USUARIOS**		**TOTAL**
	SI	**NO**	
Cree que el big data al almacenar grandes volúmenes de información puede generar información para toma de decisiones gerenciales	255	120	375
Cree que las redes sociales son un mecanismo adecuado para la captura masiva de comentarios que no se vean atados a un criterio de quien publica	227	148	375
Posee usted una cuenta en la red social Facebook	375	0	375
Cree que los servicios académicos basados en tecnologías de la información de la UTA son suficientes	9	366	375
Piensa que el uso masivo de información afecta a los sistemas informáticos	167	208	375
Conoce los servicios académicos basados en tecnología que oferta la UTA	306	69	375
TOTAL	**1339**	**911**	**2250**

Elaborado por: Investigador

Tabla 9: Frecuencias Esperadas

FRECUENCIAS ESPERADAS

PREGUNTAS	**USUARIOS**		**TOTAL**
	SI	**NO**	
Cree que el big data al almacenar grandes volúmenes de información puede generar información para toma de decisiones gerenciales	223,2	151,8	375
Cree que las redes sociales son un mecanismo adecuado para la captura masiva de comentarios que no se vean atados a un criterio de quien publica	223,2	151,8	375

Posee usted una cuenta en la red social facebook	223,2	151,8	375
Cree que los servicios académicos basados en tecnologías de la información de la UTA son suficientes	223,2	151,8	375
Piensa que el uso masivo de información afecta a los sistemas informáticos	223,2	151,8	375
Conoce los servicios académicos basados en tecnología que oferta la UTA	223,2	151,8	375
TOTAL	**1339**	**911**	**2250**

Elaborado por: Investigador

CALCULO DEL CHI CUADRADO

Tabla 10: Cálculo CHI Cuadrado

O	**E**	(O-E)	**(O-E)^2**	**(O-E)^2/E**
255	223,2	31,83	1013,36	4,54
227	223,2	3,83	14,69	0,07
375	223,2	151,83	23053,36	103,30
9	223,2	-214,17	45867,36	205,53
167	223,2	-56,17	3154,69	14,14
306	223,2	82,83	6861,36	30,75
120	151,8	-31,83	1013,36	6,67
148	151,8	-3,83	14,69	0,10
0	151,8	-151,83	23053,36	151,83
366	151,8	214,17	45867,36	302,09
208	151,8	56,17	3154,69	20,78
69	151,8	-82,83	6861,36	45,19
			CHI CUADRADO	884,98

Elaborado por: Investigador

Decisión final para 5 grados de libertad a un nivel de aceptación de 0.05 se obtiene en la tabla de Chi Tabular el valor de 11.07, y como el valor calculado del Chi cuadrado es de 884.98 se encuentra fuera de la región de aceptación, entonces se rechaza la hipótesis nula por lo que se acepta la hipótesis alternativa que dice: **“El uso de Big Data SI permite determinar la calidad de los servicios académicos de la Universidad Técnica de Ambato”.**

CAPÍTULO V

CONCLUSIONES Y RECOMENDACIONES

5.1 Conclusiones

- Al existir gran cantidad de datos en el big data resulta un trabajo muy extenso el clasificar las opiniones, además requiere expertos humanos y costos altos.

- Conforme los datos pronunciados a través de la encuesta los usuarios en su totalidad manejan las redes sociales a través de la cual generan su medio de opinión primario en varios o la mayoría de ámbitos, siendo uno de estos el académico.

- De la información levantada por medio de las encuestas a los diferentes usuarios de los servicios de TI de la UTA se puede determinar que las opiniones que son emitidas por medio de las redes sociales almacenan gran cantidad de información que ha sido opinada de forma libre y espontánea y no tiene sesgos enmarcados por una encuesta u otro instrumento similar.

- A través de las diferentes temáticas lanzadas en la red social, los usuarios pueden emitir de una manera libre su pensamiento y esto beneficia para generar una clasificación adecuada de los servicios y su calidad, de tal forma que se puede aprovechar esa información para generar una adecuada toma de decisiones.

5.2 Recomendaciones

- Tomar la gran cantidad de opiniones y estados pronunciados por los usuarios de acuerdo a un determinado tema expuesto en la red social de la UTA para viabilizar de manera adecuada las deficiencias y bondades detectadas por las opiniones emitidas.

- Generar por parte de los administradores del Sitio Oficial de la Red social, paulatinamente temas o publicaciones, que permitan generar miles de opiniones para poder seguir capturando evidencias del pensamiento libre de los usuarios.

- Automatizar el proceso de clasificación de las opiniones para cualquier tipo de publicación que se realice en Facebook, mediante una metodología para el análisis semántico del big data.

CAPÍTULO VI

LA PROPUESTA

6.1 DATOS INFORMATIVOS

6.1.1 Título: "METODOLOGIA PARA EL ANÁLISIS SEMÁNTICO DEL BIG DATA. CASO PRÁCTICO: EVALUACIÓN DE LAS OPINIONES ACERCA DE LOS SERVICIOS ACADÉMICOS DE LA UTA"

6.1.2 Institución

Ejecutora: Universidad Técnica de Ambato

6.1.3 Beneficiarios:

- Comunidad universitaria - UTA

6.1.4 Ubicación:

- **Provincia:** Tungurahua
- **Cantón:** Ambato
- **Dirección:** Av. Los Chasquis y Rio Payamino

6.1.5 Equipo Técnico Responsable

- **Investigador:** Ing. Robert Vaca A.

6.2 Antecedentes de la propuesta

En la actualidad la Universidad Técnica de Ambato posee cuentas en Facebook como uno de los medios de comunicación para con la comunidad universitaria y la ciudadanía en general, en donde no es posible determinar el nivel de veracidad de las opiniones que

se generan a partir de algún tema, en particular sobre la conformidad o no de los usuarios sobre servicios académicos brindados.

Además por el gran volumen de opiniones que se generan no es posible tabularlas para un análisis profesional que permita a las autoridades tomar decisiones acertadas sobre los servicios académicos que se está brindando a los miembros universitarios.

Se parte del supuesto de que los servicios que presta la UTA son de la más alta calidad, el criterio es subjetivo, pero puede ser medido mediante las opiniones de los estudiantes que reciben el servicio. Esta interacción es directa y además necesaria porque se recoge directamente de los beneficiarios del servicio sus opiniones. Pero esas opiniones no siempre son objetivas, o entendibles debido a la misma forma de escritura, redacción y semántica. Este tipo de opiniones almacenadas en el Big Data deberían ser susceptibles de una depuración adecuada, para que esos datos sean tratados correctamente y sirvan como soporte para la toma de decisiones a tiempo acerca de los servicios académicos que presta la UTA, y además a futuro permitan el mejoramiento continuo de los mismos. Adicionalmente, conforme se concluye en la investigación previa, la gran mayoría o casi todos los usuarios de los servicios de TI de la UTA, poseen una cuenta en las redes sociales, por lo que se hace imperativo la explotación de este medio para informar acerca de las actividades que realizan las autoridades universitarias y además emitir temas o publicaciones que generen opiniones, de las cuales se pueden hacer estudios que dirijan el trabajo de las mismas, a través de mejorar procesos, servicios, eliminar unos y mejorar otros así como incrementarlos.

6.3 Justificación

La presente propuesta se justifica para lograr la automatización de la clasificación de las opiniones y reacciones acerca de la calidad de los servicios académicos que oferta la UTA, siendo el alcance de la metodología propuesta no solamente limitado para servicios académicos sino también para cualquier tipo de eje de desarrollo que requiera opiniones masivas que generalmente se originan en páginas de Facebook, en donde la espontaneidad de las mismas contienen demasiada subjetividad en su redacción, en la

que es muy difícil determinar la intensión del individuo que opina, dificultando tener claridad sobre los resultados.

La metodología propuesta elimina la subjetividad por lo tanto se concluye que las opiniones filtradas son adecuadas para clasificar a los servicios ofertados. La metodología planteada en este trabajo muestra que es posible implementar un proceso que permite agrupar los servicios en buenos y mejorables. En este sentido y dado que se ha mostrado la factibilidad de la misma se propone su validación en su caso de estudio específico acerca de uno de los servicios académicos de la UTA a partir de las opiniones de los estudiantes.

La solución al problema planteado radica en la utilización de un conjunto de aplicaciones informáticas que permiten analizar las opiniones y brindar información válida que servirá para toma de decisiones a nivel gerencial. El mejoramiento continuo de los servicios es un objetivo primordial de cualquier institución educativa y de allí, que con la aplicación de la propuesta se persigue obtener de forma directa el criterio de los usuarios que en este caso son los estudiantes. Si conocemos de forma adecuada que es lo que piensan los estudiantes de un servicio, se pueden plantear las mejoras adecuadas o concluir que el servicio tal como está planteado ofrece lo necesario para que el estudiante pueda realizar un proceso y/o obtener la información correcta

6.4 Objetivos

Objetivo General

Desarrollar una metodología para automatizar el análisis semántico de las opiniones en el big data y determinar la calidad de los servicios académicos de la UTA.

Objetivos Específicos

- Utilizar herramientas tecnológicas para obtener y analizar las opiniones sobre servicios académicos de la UTA.
- Aplicar la metodología en un caso de estudio real con el servicio de matriculación de los estudiantes de la UTA

6.5 Análisis de Factibilidad

La propuesta es factible por cuanto se dispone de las herramientas tecnológicas para su aplicación, así como también las opiniones de los estudiantes en el sitio Facebook oficial de la UTA, sobre publicaciones relacionadas a servicios académicos.

6.5.1 Factibilidad Técnica

Es factible técnicamente por cuanto se dispone de la tecnología necesaria (facebook, R, PMI, hueca) para la aplicación de la metodología en la propuesta presentada.

6.5.2 Factibilidad Organizacional

Existe la disponibilidad de la información, así como la generación de nuevas opiniones sobre servicios académicos que oferta la UTA, y al ser el tema propuesto una herramienta para toma de decisiones a nivel académico, se garantiza su ejecución y aplicación por parte de la institución

6.5.3 Factibilidad Económica

Al utilizar herramientas tecnológicas libres que no requieren ningún comprometimiento presupuestario, la propuesta es económicamente factible.

6.6 Fundamentaciones

6.6.1 Filosófica

Para realizar el proyecto se utilizó el paradigma filosófico crítico propositivo ya que se toman las opiniones vertidas en Facebook de la UTA y se plantea una solución al problema de la gran cantidad de información subjetiva.

El conocimiento sobre los algoritmos permite diseñar el sistema y hacer uso de la metodología planteada en este proyecto de investigación, y es importante dar a conocer los conceptos fundamentales que se utilizan para el desarrollo del proyecto.

Definiciones Generales

Redes Sociales

Las redes sociales se definen como un conjunto bien delimitado de actores, individuos, grupos, organizaciones, comunidades, sociedades; que para el caso de la Universidad Técnica de Ambato sus actores principales son la institución, estudiantes, docentes, administrativos y comunidad en general, vinculados entre sí por un objetivo en común relacionado a la gestión universitaria y al proceso enseñanza – aprendizaje. Las características de estas relaciones se usan para interpretar los comportamientos sociales de las personas implicadas y conocer sus opiniones sobre temas de interés común.

Facebook – Big Data

Este sistema busca convertir los sistemas analíticos convencionales en información simple y sencilla que facilite la toma de decisiones en tiempo real por medio de infraestructuras tecnologías y servicios que han sido creados para dar solución al procesamiento de enormes conjuntos de datos estructurados, no estructurados o semi-estructurados. Diferentes organizaciones han tratado de solucionar este problema pero se han estancado en el proceso, siendo pocos los que han emergido un claro ejemplo es Hadoop que es una plataforma de código abierta capaz de analizar grandes cantidades de información, este inspirado en el proyecto de Google File System (GFS) el cual consiste en dividir en dos tareas (mapper – reducer) para manipular los datos distribuidos a nodos de un clúster logrando un alto paralelismo en el procesamiento. Hadoop el mismo que está compuesto de Hadoop Distributed File System, Hadoop Map Reduce (Barranco, 2012)

¿De dónde provienen todos estos datos?

Todos los días se producen miles de millones de datos, los cuales son generados por las personas (e-mail, publicaciones, historiales, etc.), transacciones de datos (llamadas, facturación, transacciones, etc.), E-marketing y web (publicidad y demás elementos

relacionados al marketing), Machine to machine (GPS, Sensores de temperatura, luz, etc.) y Biométrica (Información de seguridad, defensa y servicios de inteligencia).

En la Universidad Técnica de Ambato el big data se conforma por los miles de comentarios que se reciben todos los días mediante el sitio oficial de facebook, relacionado a publicaciones realizadas sobre eventos o servicios que la UTA brinda a estudiantes, docentes, personal administrativo y comunidad en general.

Características de Big Data

1) Volumen de los datos: una gran cantidad de información producida en muy poco tiempo. Ésta se genera a través de facebook de la UTA.

2) Velocidad con la que se generan los datos: ya se mencionaba la rapidez con que se genera la información: 5 exabytes cada 10 minutos.

3) Variedad de los datos: existen dos clases de datos. Estructurados: bases de datos organizadas y divididas de forma lógica. No Estructurados: fotografías, videos, posts, grabaciones de audio, para el desarrollo de la propuesta de utilizan posts de facebook de la Universidad.

4) Valor: qué parte de esos 5 exabytes es valiosa. Información que se obtiene depurando las opiniones, descartando la subjetividad de la misma.

5) Veracidad de los datos: no toda la información generada es confiable, incluso hay muchas mentiras generadas todos los días. Es necesario analizar los datos y determinar cuál es confiable y cuál es incorrecta. (Malacara, 2014).

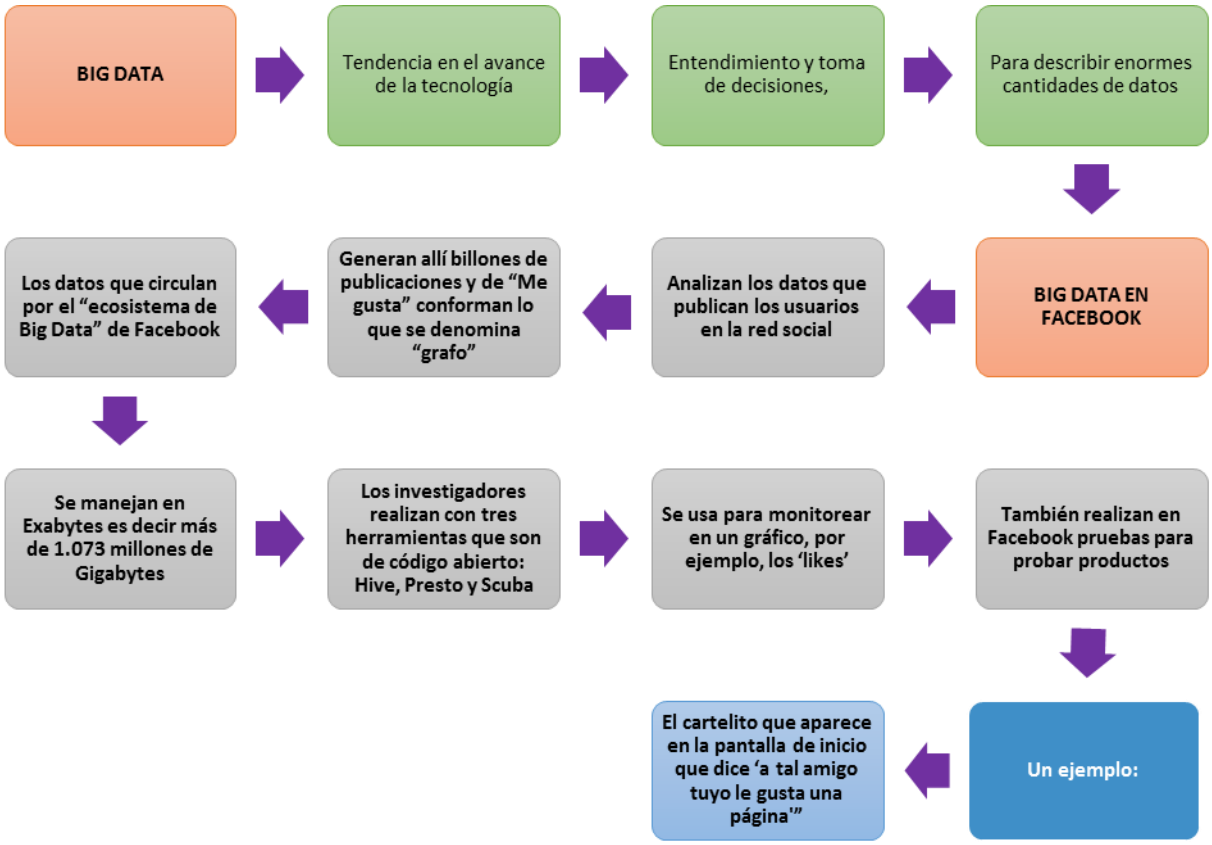


Figura 8: Big Data y Facebook
Fuente**:** Malacara, 2014

Facebook - Post

Es todo contenido que aparece en tu muro, los mismos que pueden contener actitudes positivas o negativas incluidas en texto o imágenes que transmiten información interesante o novedosa, lo importante es que esta información publicada contará con diversas opiniones de aceptación o rechazo de los usuarios interesados en tus mensajes.

Facebook UTA– Estructura Post

Figura 9: Facebook UTA Estructura post
Elaborado por: Investigador

Facebook Api_Graph

La API Graph es una macro que permite extraer datos desde la página de Facebook, pudiendo ser estos: comentarios, estados, likes, entre otros. Está basada en HTTP y necesita de inicio de sesión en facebook para su utilización.

Facebook ID

Cada página creada en facebook tiene su número personal de identificación otorgado para caracterizar de manera única a la página. Dicho identificador puede ser obtenido mediante herramientas online.
Para obtener el Facebook ID del sitio oficial de la UTA ingresamos a https://findmyfbid.com/ y obtuvimos el siguiente resultado: 1431460530422719

JSON (JavaScript Object Notation)

Es un formato para intercambiar datos y se caracteriza por que puede ser leído por cualquier lenguaje de programación, y por lo tanto puede ser utilizado para intercambio de información entre distintas tecnologías.

Los comentarios obtenidos mediante el explorador de la API Graph de facebook tienen una estructura JSON.

Ejemplo: comentarios obtenidos de una publicación en el sitio oficial de la UTA

Figura 10: Comentarios en JSON
Elaborado por: Investigador

Punto de información mutua: *PMI*

PMI se basa en el cálculo de la intersección de términos de una frase, ésta intersección permite obtener un indicador de la participación de los mismos en una frase, una vez que se obtiene este indicador se puede determinar si la frase puede ser clasificada como Excellent(Positiva) o Poor(Negative) por las personas, para ello:

$$SO(phrase) = PMI(phrase, \text{“excellent”}) - PMI(phrase, \text{“poor”})$$

La propuesta de Bing[21], propone que se puede obtener la orientación semántica de la frase (clasificación de la frase) por medio de restar el indicador de cuando la frase fue clasificada como positiva o negativa por las personas.

Elementos del archivo PMI

Librería tm

Contiene las utilidades necesarias para limpiar los datos. Además posee el diccionario con el cual se relaciona cada palabra, ejemplo la palabra "good" y su porcentaje de positividad y de negatividad.

Librería RWeka

Permite llamar desde el lenguaje R a Weka para ejecutar sus algoritmos, para nuestro caso para ejecutar el Kmeans.

Librería caret

Permite graficar los resultados obtenidos mediante el lenguaje R en relación al algoritmo Kmeans de Weka.

Weka

Significado.- La Weka (Gallirallusaustralis) es un ave endémica de Nueva Zelanda. Esta Gallinácea en peligro de extinción es famosa por su curiosidad y agresividad. De aspecto pardo y tamaño similar a una gallina las wekas se alimentan fundamentalmente de insectos y frutos.

Esta ave da nombre a una extensa colección de algoritmos de Máquinas de conocimiento desarrollados por la universidad de Waikato (Nueva Zelanda) implementados en Java [1, 2]; útiles para ser aplicados sobre datos mediante las interfaces que ofrece o para embeberlos dentro de cualquier aplicación. Además Weka es un programa de libre distribución y difusión contiene las herramientas necesarias para realizar transformaciones sobre los datos, tareas de clasificación, regresión, clustering, asociación y visualización. Weka está diseñado como una herramienta orientada a la extensibilidad por lo que añadir nuevas funcionalidades es una tarea sencilla.(Morate, 2000)

Algoritmo Kmedias

K-Vecinos más cercanos o KNN es un método tanto de predicción como de clasificación. El pronóstico sobre una nueva observación se basa en el pronóstico no en todas las observaciones disponibles, sino en las observaciones que se parecen más a la nueva observación. Kmeans es un método de agrupamiento o *cluster*, tiene el objetivo de tratar de encontrar en la base de datos grupos de observaciones que tienen características semejantes. Si los clusters encontrados tienen buena calidad tiene que ser diferente a lo encontrado para los demás grupos. En agrupamiento se trata de maximizar la variación que existe intercluster y al mismo tiempo disminuir la variación extra cluster. La distancia de cluster a cluster es grande se llama distancia entre inter cluster, pero en el mismo cluster tiene que ser mínima.

Las técnicas de cluster no jerárquico necesitan fijar de antemano el número de conglomerados en los que se necesitan agrupar los datos. El objetivo es intentar obtener una clasificación por grupos en el sentido que la dispersión dentro de cada grupo que se forme sea la menor posible. La dispersión dentro de un grupo se mide en términos de las observaciones al centroide del grupo. La separación entre grupos sea lo suficientemente grande, puesto que esta se mide como la distancia que separa los centroides de los grupos. Si se refiere a la distancia entre grupos como la distancia que separan los diferentes centroides. Este criterio de la varianza es necesario para buscar la solución cluster. El más conocido es algoritmo de Kmedias. Este algoritmo comienza con la configuración al azar de grupos con su correspondiente centroide eligiendo un primer individuo u observación, y asignando posteriormente los otros casos según el efecto que esta tenga sobre la dispersión de los grupos, es decir que si tenemos un grupo de observaciones en este método se agregan las mismas a aquellos grupos en los que no se vea modificada de forma significativa la dispersión del mismo.

El valor mínimo de varianza determina una configuración de nuevos grupos con sus respectivas medias y se deben asignar otra vez todos los casos a estos nuevos centroides, este proceso se repite hasta que ninguna transferencia pueda disminuir más la varianza intra-grupos, o se alcance otro criterio de parada como puede ser un número

limitado de iteraciones, o simplemente que la diferencia obtenida entre los centroides de dos pasos consecutivos sea menor que un valor prefijado. En este sentido los centroides se recolocan mejorando el agrupamiento. La diferencia obtenida entre los centroides es menor que un valor prefijado, es decir que los centroides no se mueven mas allá de una determinada distancia considerando que la solución cluster es estable (Nguyen, Hoang, Van, Van, & Duy, 2016).

Suponiendo que existen tres variables con 3 observaciones o individuos. Con k = 3, implica que las observaciones estarán formadas por las observaciones 2,5,6 el cluster 2 por la 1,3 y el cluster 3 por la observación 4.

	X1	X2	X3
Obs1	5	9	20
Obs2	6	11	2
Obs3	4	5	20
Obs4	6	9	46
Obs5	5	7	1
Obs6	3	1	12

Paso 1:

Se parte de k clusters iniciales:

Cluster 1 ={2.5.6}

Cluster 2={1,3}

Cluster 3 = {4}

Paso 2:

Se calculan los centroides de cada cluster:

CentroideCLuster 1({2,5,6})= (14/3, 19/3,15/3)

Coordenada para X1: (6+5+3)/3 = 14/3

Coordenada para X2: (11 + 7 + 1)/3 = 19/3

Coordenada para X3: (2 + 1 + 12)/3 = 15/3

CentroideCluster 2 ({1,3}) = (9/2,7,20)

CentroideCluster 3({4}) = (6,9,46)

Paso 3

Se calcula la suma de varianzas dentro de cada grupo. En este caso 136.805

Paso 4:

Se comprueba si la reasignación de una observación a otro cluster implica una reducción en la suma de varianzas dentro de cada grupo. En este caso es cuando se mueve la observación 6 al cluster 2. Esto implica que la varianza llega a 85.6 que es menor. Por lo tanto como ha habido un cambio de las observaciones, y es necesario calcular nuevamente los centroides.

Paso 5

Cluster 1 = {2,5} CentroideCluster 1 = (11/2,9,3/2)

Cluster 2= {1,3,6} CentroideCluster 2 = {4,5,52/3)

Cluster 3 = {4} CentroideCluster 3 = {6,9,46}

Paso 6

Se comprueba si algún otro movimiento reduce aún más la suma de varianzas y se repite el procedimiento anterior hasta que no se consiga ninguna mejora. En este caso se comprueba que no hay ningún movimiento que haga disminuir las varianzas y por lo tanto, la solución final es la propuesta en el Paso 5.

Para el caso de esta tesis, las observaciones se corresponden con los datos de la orientación semántica y la polaridad del texto, y se procede a obtener de la misma manera como se ha descrito en los pasos para los grupos.

Gráfico Explicación Kmeans

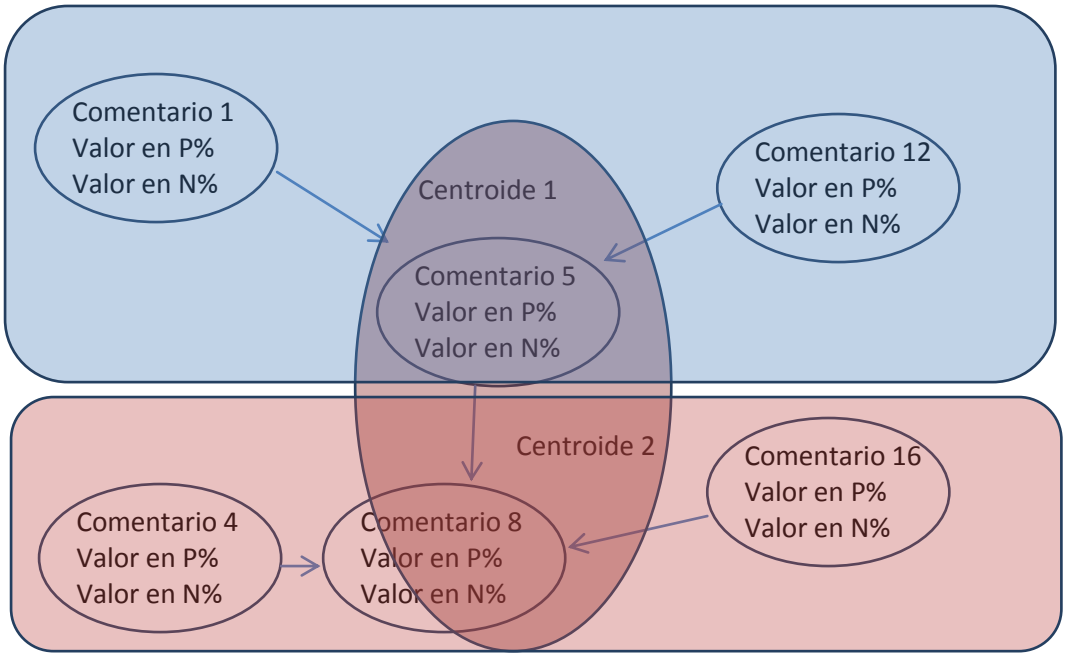


Gráfico 11. Gráfico Kmeans

Elaborado por: Investigador

Análisis de sentimientos

Métodos para identificar emociones y sentimientos en texto

Levallois, C (2013) propone desarrollar un motor de detección diseñado para sentimientos positivo, negativo o neutro en tweets, como partes principales consta de la detección de rasgos semánticos del tweet(emoticones y onomatopeyas), evaluación de hashtags; una lista de n-gramas de la descomposición de los tweets, los emoticones y onomatopeyas tienen fuertes indicios de sentimiento, pero así mismo una ortografía variada, utilizan las exclamaciones más comunes para capturar la variedad de formas que pueden asumir (disponible en: www.umigon.com). Para la evaluación de los hashtags aplican una serie de heurísticas. En la descomposición del tweet hay una lista

de n-gramas (unigramas, bigramas y trigramas) en donde son recorridos en cada tweet y realizan comprobaciones de su presencia en diccionarios de términos léxicos. Si se encuentra en el diccionario de cualquier n-grama se aplica la heurística, para tener como resultado una clasificación (positivo, negativo o neutro), para el análisis de sentimientos utilizaron cuatro diccionarios: positivo, negativo, fuerza del sentimiento y negaciones; los cuales fueron creados manualmente. Este sistema implementado por el autor del paper obtuvo una precisión promedio (positivo y negativo) de 69.02%.

Hangya (2013) desarrolló un sistema para la normalización de los tweets, en su implementación utilizaron las herramientas MALLET basado en Java para el procesamiento de lenguaje natural, el sistema realiza tareas donde toma todas las palabras y las convierte a minúscula (utilizando el algoritmo de PorterStemming), sustituyen @ y # por notaciones [usuario] y [tag], los emoticones se clasifican en positivos y negativos, se eliminan caracteres innecesarios, en el caso de existir repeticiones de palabras en cada carácter reducen la longitud y como punto final realizan un filtrado de palabras que no tiene significado. Después de realizar la normalización de los mensajes determinan la polaridad de cada palabra en donde utilizan el diccionario de sentimientos SentiWordNet, donde consideran una palabra como positiva si el valor positivo es mayor a 0.3; como negativa si es mayor a 0.2 y como neutral si es mayor a 0.8. Luego de calcular la polaridad se basan en las tres características principales para cada tweet: como el número de palabras positivas, negativas y objetivas respectivamente; luego verifican si a una palabra positiva le precede una negación, si es así la polaridad se invierte. Cada decir que como clasificador de aprendizaje utilizo máxima entropía dando como mejor resultado una precisión del 54%, usando un modelo basado en la obtención de características y la normalización de los tweets.

Dubiau (2013) Propone analizar y comparar técnicas de procesamiento de lenguaje natural para la clasificación de documentos a partir de la identificación y extracción de información subjetiva como opiniones y sentimientos. Para la implementación las herramientas y frameworks utilizados para la clasificación subjetiva de texto son (NLTK, MEGAM, Sci-Kit Learn, Freeling), el análisis de sentimientos busca detectar

estas emociones: quién las posee (titular); y cuál es el aspecto que genera la emoción (target). Para ello utilizan un tamaño de corpus entre 500 y 22.000 documentos del conjunto de datos de google play y guía óleo, los cuales son sitios de críticas gastronómicas(disponible en www.guiaoleo.com), en el cual dichos sitios emiten los usuarios opiniones sobre restaurantes de la ciudad de Buenos Aires y tiene una calificación en las categorías de comida, ambiente y servicio, estas categorías son calificadas con un puntaje ya sea malo regular bueno o muy bueno, el criterio de asignación de etiquetas a los comentarios de los usuarios fue dar como Positivo a comentarios mencionados que tenga como categoría 10 o superior, así mismo para dar la etiqueta a un comentario Negativo si se encuentra en la categoría COMIDA un valor de 1 o 2. Construyen un conjunto de entrenamiento y otro de prueba, estos se encuentran en un formato de Json, obteniendo un tratamiento de stemming, lematización, tokenización y n-gramas. La selección de atributos o características que utiliza para métodos supervisados es la presencia y frecuencia de unigramas, adjetivos y combinaciones de ellas, y de esa manera sacar la efectividad de pre procesamiento en el análisis de sentimiento. Los corpus que utilizan se encuentran balanceados y desbalanceados porque de esa manera puede analizar el comportamiento de los métodos en estudio cuando el conjunto de datos se encuentra fuertemente desbalanceado y luego comparar los resultados con los obtenidos para corpus balanceados. El mejor resultado dado por (Dubiau, 2013) fue Redes bayesianas Multinomial aunque utilizo otros modelos que han implementado para la comparación como son Máxima Entropía, Maquina de Soporte Vectorial, Árbol de decisión y Turney dichos métodos determinan un estado de efectividad, y luego estos valores los comparan gráficamente para ver cuales tienen un valor de efectividad más acertado. Por lo tanto consideró el modelo de Redes Bayesianas Multinomial porque permite tomar todas las características de todos los términos del vocabulario del corpus de entrenamiento teniendo en cuenta su frecuencia de aparición. Es por ello que al analizar dichos modelos en su experimentación el que tiene menor margen de error son Redes Bayesianas con un 8%, MaxEnt con el 6% y SVM con el 7%, y su valor de efectividad en cambio en NB 98%, MaxEnt 94% y SVM 93% escogiendo para pruebas de experimentación las Redes Bayesianas por su fácil entrenamiento, simplificando la conversación de los datos

textuales a datos estadísticos de frecuencia de la polaridad positiva y negativo en un corpus determinado.

Es así que la investigación desarrollada y detallada anteriormente trabaja con textos no estructurados siendo la técnica apropiada y una base para la construcción de la metodología planteada para la EVALUACIÓN DE LAS OPINIONES ACERCA DE LOS SERVICIOS ACADÉMICOS DE LA UTA

Polaridad – Definición

A través del análisis semántico de documentos se pueden asignar categorías, se aborda la carga emotiva si la misma tiene una carga emocional positiva, negativa o neutra, para ello, se analiza con algoritmos de minería de texto los párrafos de un texto y se los clasifica según su carga emotiva, lo cual se conoce como polaridad

El resultado es iniciar un análisis emocional de los documentos publicitarios, como también las sentencias, los mismos que puedan ser utilizados para conformar estructuras semánticas y ontológicas que permitan compartir modelos de emociones positivas y con ello compartir conocimientos. Se identifica la polaridad del texto (en red social me gusta para positivo y no me gusta para negativo) para luego tratar de interpretar esto como parte de una emoción positiva o negativa. No se puede asegurar la polaridad de un texto, pues inclusive este depende del estado de ánimo de la persona, de allí que los resultados sean solamente una guía para la toma de decisiones de parte de un experto humano en un determinado espacio comunicacional o jurídico como nueva forma de comunicación persuasiva.

En minería de texto se trabaja con correspondencia semántica de las opiniones, una vez que las mismas fueron filtrada mediante la orientación semántica. Para ello es necesario el uso de Los diccionarios que se describen a continuación, como:

- ElhPolar: relaciona con la palabra con el nivel de polaridad. (Por ejemplo: a_ciegas (negativo) (Saralegi & San Vicente, 2013), este diccionario está disponible en español, con el mismo podemos emular las propuestas de Poria con SenticNet(Poria, Gelbukh, Cambria, Hussain, & Guang-Bin, 2014).

- EMS relaciona palabras con emociones. (Por ejemplo, 'victimizado': 'traición'), este diccionario está disponible en Inglés, con el fin de proveer ejemplos más concretos, las palabras han sido traducidas al español[5]. Con este diccionario en particular, podemos emular las propuestas de Poria(Poria, Gelbukh, Cambria, Hussain, & Guang-Bin, 2014), así como con WNA (Espacios afectivos).

Minería de Texto

Cobo, A. R., & Martínez, M. (2009) Proponen el uso combinado de metodologías de minería de texto y técnicas de inteligencia artificial con el fin de optimizar los mecanismos de categorización, extracción automática de conocimiento y la agrupación de colecciones documentales, es decir hace mención a un modelo de gestión documental integral para el proceso de información no estructurada, el modelo que implementan es una aplicación de uso intuitivo, multilingüe que integra técnicas de minería de texto. Abordan tres problemas importantes de la implementación de técnicas de minería de texto, la extracción de documentos, recuperación de información, y la categorización de documentos en la que se asigna a cada documento una o varias categorías. La clasificación se la puede realizar mediante categorización o clustering, en el primer caso se habla de clasificación supervisada, mientras que en segundo caso se utiliza el concepto de aprendizaje no supervisado. El autor de este paper utiliza un modelo vectorial que permite la representación de documentos a partir de un vector de pesos o palabras que se encuentran en el texto, para luego realizar la eliminación de palabras que no tiene valor significativo, luego se realiza el pre procesamiento de texto para que la información se encuentre con una estructura igual, y poder mostrar la información y extraer datos en forma no estructurada. Para la estructura de la información utiliza una bolsa de palabras o llamado también lista de palabras en este caso la bolsa de palabras es estructurada en varios idiomas, una vez que han estructurado la información con más

[5]EMS can be obtained from: https://pythonism.wordpress.com/2013/06/16/elementary-sentiment-analysis-on-a-text-using-python/#more-1256.

relevancia, propone que la información nueva debe ser presentada mediante una interfaz de comunicación con el usuario.

Para la implementación del modelo optaron por tecnologías de código abierto, tomando como núcleo la aplicación OWL el cual incorpora funcionalidades básicas de un sistema de gestión documental, el sistema de gestión documental propuesto por los autores de este paper funciona mediante protocolos de comunicación para el acceso a los datos por parte del usuario, en cuanto a la minería de texto en la parte central del sistema de gestión documental se han implementado la extracción automática de conocimiento, utilizando recursos lingüísticos como glosarios para el inicio del análisis, luego de contar con la lista de términos, que determinan la forma de cada palabra, llamado también análisis morfológico del texto extraído para identificar sustantivos, adjetivos y verbos.

Estos autores como resultados experimentales del uso del modelo y la aplicación en el proceso de clasificación utilizo una colección de 250 documentos científicos asociados a 5 categorías diferentes. Es así que para aumentar la complejidad del proceso de clasificación seleccionaron 125 documentos escritos en idioma inglés y el restante de la colección de 250 en español, donde ven que la minerías de texto combinada con modelos de optimización ayudan a una adecuada gestión de grandes volúmenes de información no estructurada que se genera en el contexto de las organizaciones.

Programa R

Coeficiente de Kappa.- El **Coeficiente kappa** de Cohen es una medida estadística que ajusta el efecto del azar en la proporción de la concordancia observada para elementos cualitativos (variables categóricas).

Matriz de confusión.- Es una herramienta estándar de evaluación de modelos estadísticos, ordena todos los casos del modelo en categorías, determinando si el valor de predicción coincide con el valor real. El gráfico se crea comparando los valores reales con los valores de predicción para cada estado de predicción especificado. Las filas de la matriz representan los valores de predicción para el modelo, mientras que las columnas representan los valores reales. (Uriarte, 2003)

6.7 Metodología. Modelo Operativo.

Se basa en reglas obtenidas por medio de la agrupación de los servicios, ésta agrupación genera valores medios a través de algoritmos informáticos.

- Utilizar los comentarios obtenidos en el fichero de entrenamiento
- Someter estos comentarios a análisis de polaridad y de orientación semántica
- Aplicar el algoritmo de clusterización con el fín de analizar que tipo de servicios se pueden considerar como positivos y cuales son aquellos que en para los usuarios de esta página de Facebook se consideran como negativos.

Construcción de la metodología para automatizar el análisis semántico de las opiniones en el big data y determinar la calidad de los servicios académicos de la UTA.

Procesos

Para la construcción de la metodología se plantea lo siguiente:

1. Selección de página y obtención de comentarios
2. Filtrado de las opiniones
3. Análisis de sentimientos
4. Resultados de la Clasificación

1. Selección de página y obtención de comentarios

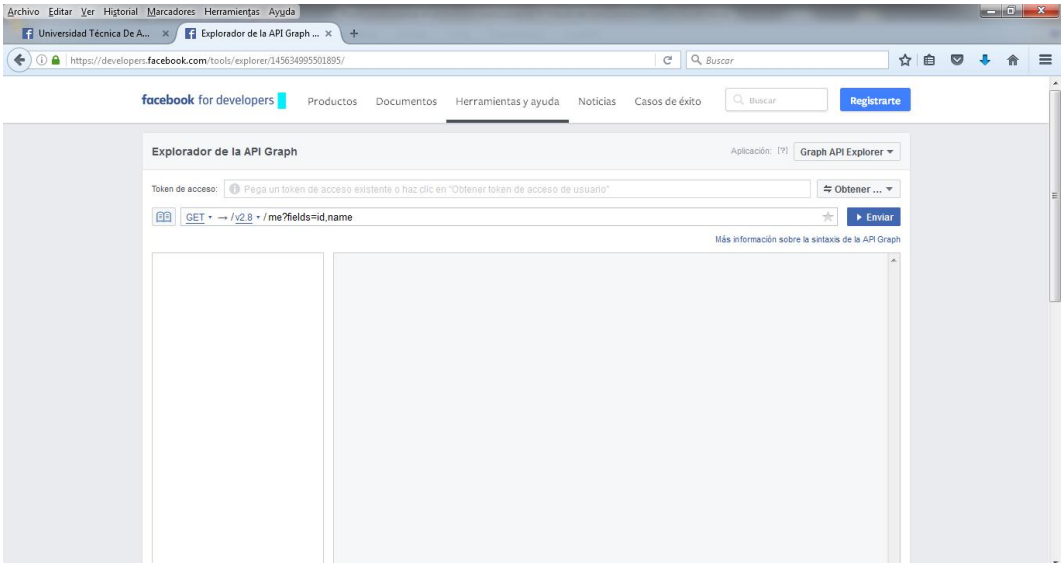
El administrador del sistema se convierte en el principal actor del proceso, es quien junto a las autoridades selecciona el campo de acción (temática a proponer) en Facebook con el fin de que la comunidad comente. Estos comentarios son la base fundamental para el analisis de opiniones y la clasificación de los servicios.

a) **Selección de la temática.-** Se debe ingresar al sitio oficial deseado y seleccionar la temática

Figura 11: Selección de la temática
Elaborado por: Investigador

Se propone analizar un servicio específico.

b) **Obtención de comentarios.-** Para ello se utiliza el Api_Graph de Facebook: https://developers.facebook.com/tools/explorer/ e iniciar sesión con la cuenta de Facebook



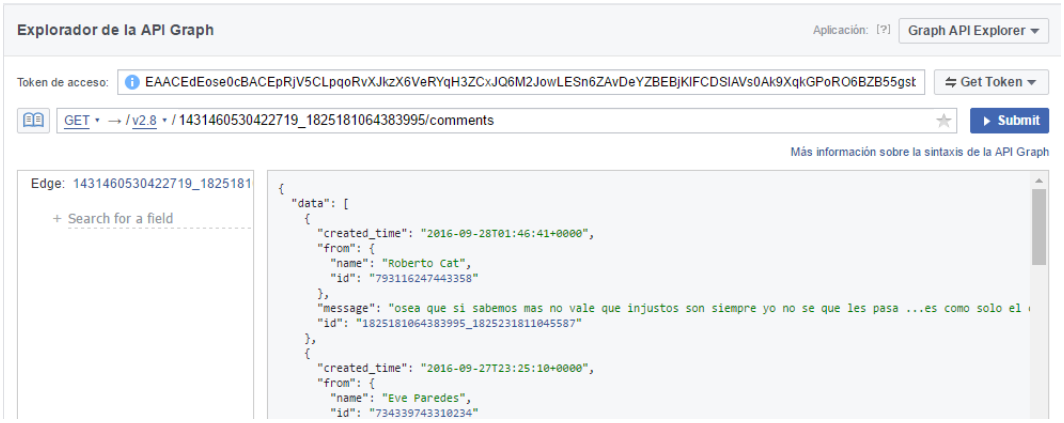


Figura 12:. Comentarios
Elaborado por: Investigador

La numeración: 1431460530422719, se corresponde con el identificador del tema. Este numero se obtiene ingresando a http://findmyfbid.com/

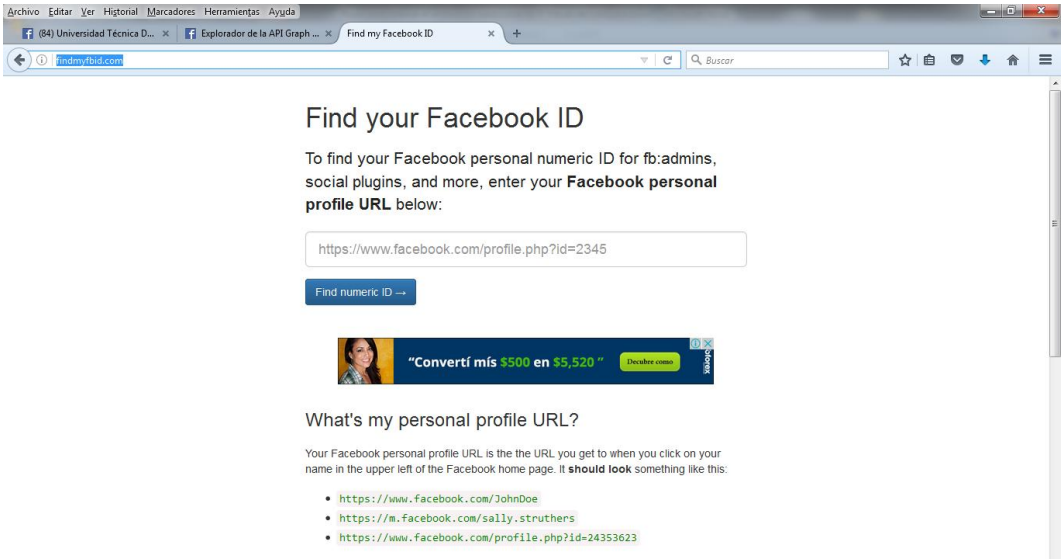


Figura 13: Facebook ID
Elaborado por: Investigador

Y en el lugar que solicita dirección se pone la dirección de Facebook seleccionada para nuestro caso: https://www.facebook.com/UniversidadTecnicadeAmbatoOficial/

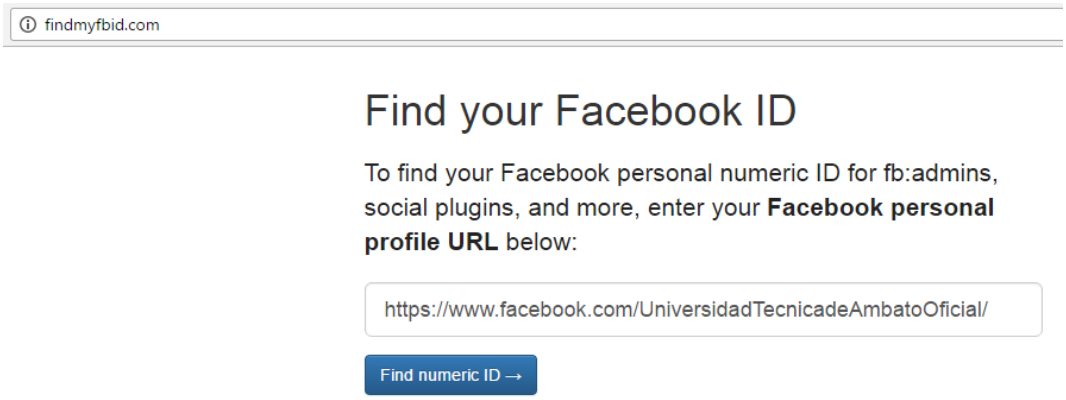


Figura 14: Identificador del tema
Elaborado por: Investigador

Luego al presionar "Find numeric ID" da como resultado un valor numérico extenso Luego se busca una opinión en el sitio facebook y se presiona sobre la fecha y se copia el número que se genera después de "_id=".

Figura 15: Identificador de comentarios
Elaborado por: Investigador

Al ejecutar todo el proceso, se obtiene los comentarios:

- Pulsar en “Obtener” y “Obtener token de acceso de usuario”.
- Por último clic en enviar y se genera el documento con las opiniones

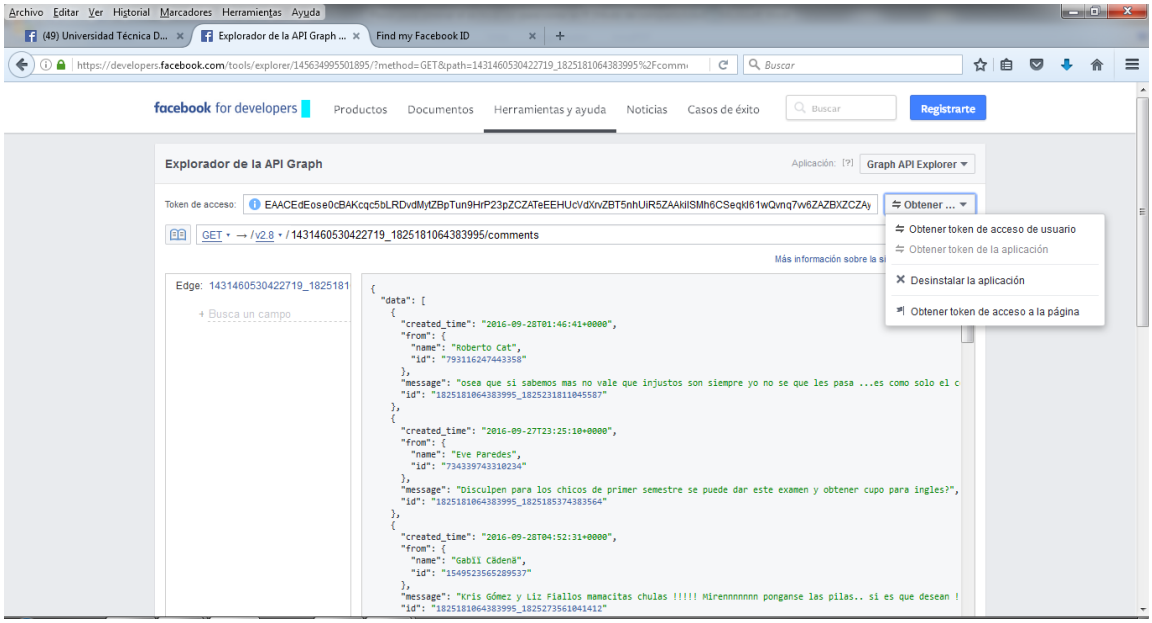


Figura 16: Obtención de todos los comentarios
Elaborado por: Investigador

Exportación a CSV

- Se utiliza la aplicación Konklone:https://konklone.io/json/
- Pegar los comentarios obtenidos en api de facebook (Gráfico 22)
- Presionar en “Downloadtheentire CSV” para obtener el archivo csv.

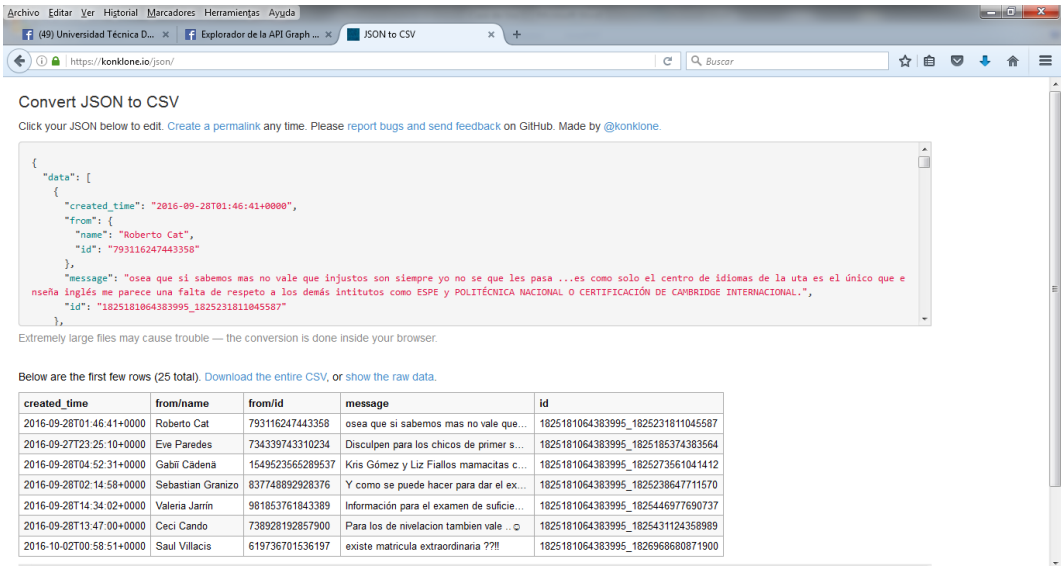


Figura 17: Aplicación Konklone
Elaborado por: Investigador

Las opiniones del CSV deben ser filtradas por el Administrador de la aplicación antes de proceder a ejecutar la polaridad y el analisis de sentimientos previo a la clasificación por clusters de los servicios.

2. Filtrado de las opiniones

No todas las opiniones que se vierten para el tema son válidas, esto implica que el administrador ayudado por un experto en el tema, seleccione aquellas opiniones que van a formar parte del conjunto que serán analizadas mediante PMI y el análisis de sentimientos.

3. Análisis mediante PMI

a) Utilización del programa R

- En R se abre el archivo PMI.R (archivo elaborado y adjunto a la propuesta)
- Se cambia el path de donde están los comentarios en csv y en “Editar” se selecciona “Ejecutar Todo”, entonces aparecen errores de librerías o funciones
- Entonces en paquetes – instalar paquetes, seleccionamos el idioma y buscamos en la lista las funciones o paquetes a instalar (tm, RWeka, caret)

Al cargar el Archivo PMI.R, este contiene los comandos para ejecutar en R en donde las “D” corresponden al número de archivos “POSITIVOS” y las “R” al número de archivos “NEGATIVOS” a ejecutar. En el archivo PMI.R adjunto a la propuesta existen 71 D, es decir 71 archivos, entonces dentro del path se debe copiar 71 archivos con comentarios..

Se recomienda bajar del sitio de Facebook, el número de comentarios equitativamente por los servicios a analizar. Ejm Servicio1 (35 archivos), servicio2 (35 comentarios)

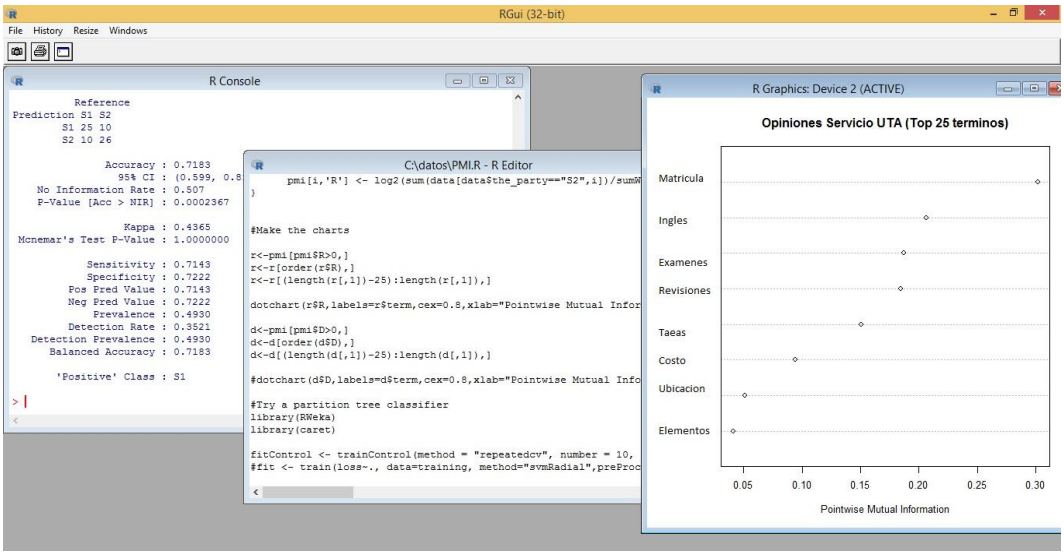


Figura 18: Primeros resultados
Elaborado por: Investigador

4. Resultados de la Clasificación

En la siguiente figura se observa los resultados, matriz de confusión y agrupamiento:

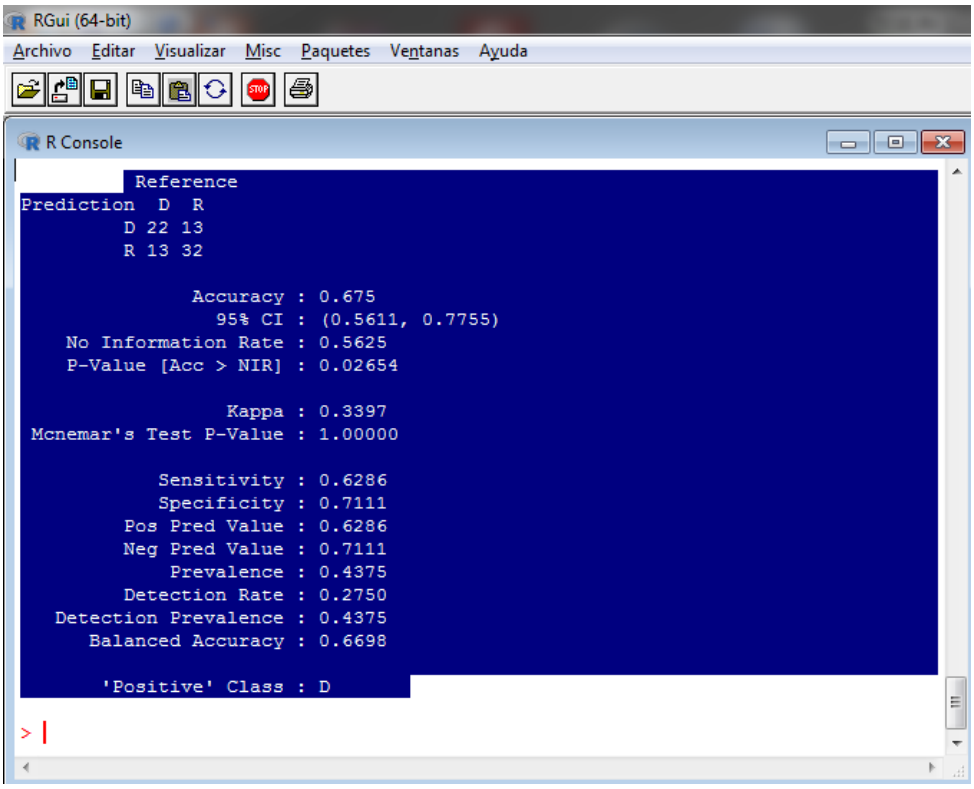


Figura 19: Resultados de la clasificación
Elaborado por: Investigador

La matriz de confusión relacionada con referencia y predicción, nos indica que existen 22 comentarios positivos que realmente eran positivos desde el inicio. Hay que tomar en cuenta aquí, que en un mismo párrafo según lo que se haya escrito existen frases positivas y negativas por ende el sistema debe responder y decidir sobre si es más positivo o negativo. Existen 13 considerados positivos pero realmente son negativos, a esto se refiere como error de predicción del sistema es decir falsos positivos. Existen 32 que son comentarios negativos y en efecto desde su inicio son clasificados por el experto como tal.

La precisión del sistema esta medida por Accuracy que en este caso es cercano a 0.70 (un valor aceptable) y por sensitividad y especificidad que indica que el sistema es capaz de identificar con claridad a los falsos positivos y a los falsos negativos (0.71). La capacidad de predicción supera el 60% para los valores positivos y negativos lo que indica que el sistema es aceptable. La tasa de detección y la prevalencia están indicando posibles errores del sistema, se observa una sensibilidad hacia los falsos positivos del

62% por ende la prevalencia esta indicando un 43% de error. Además se calculo en balance del sistema la cual esta en un 66% por ende se puede decir que el resultado parte de un balance inicial del sistema y por lo tanto se pueden aceptar los mismos. Existe un índice de Kappa (índice de observación en el experimento) cuyo valor es del 33%. En este caso al ser comentarios libres, se puede tener distintas apreciaciones sobre lo que tratan de decir las personas cuando escriben algo. Con ese valor se está diciendo que existe una coincidencia mayor al 65% en la observación de la polaridad para cada párrafo o archivo en proceso, esto relacionado a los P-value que hablan de un error del 32% también para las observaciones y coincidencias del sistema.

Proceso de la Metodología

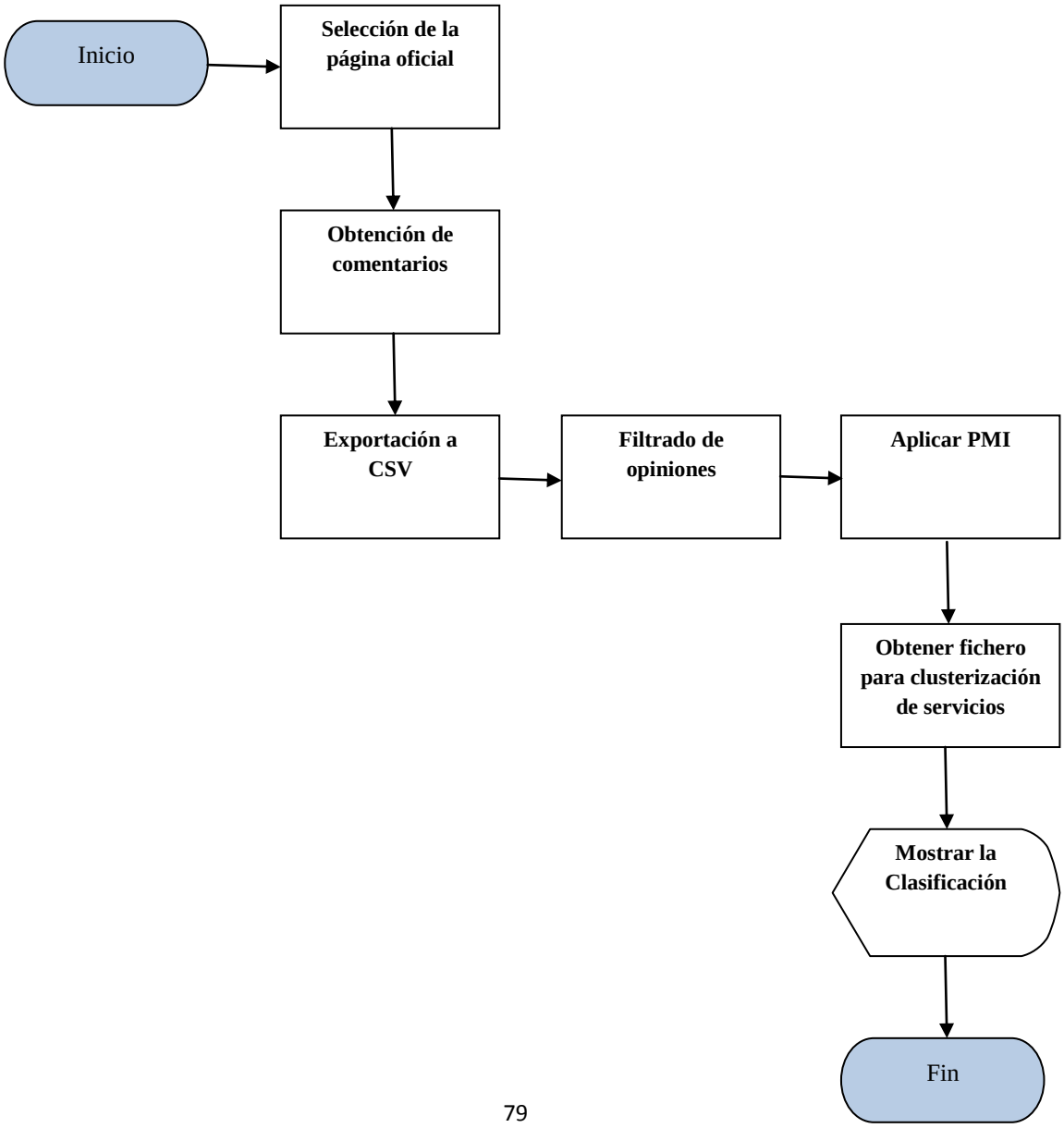


Figura 20: Diagrama de procesos de la metodología
Elaborado por: Investigador

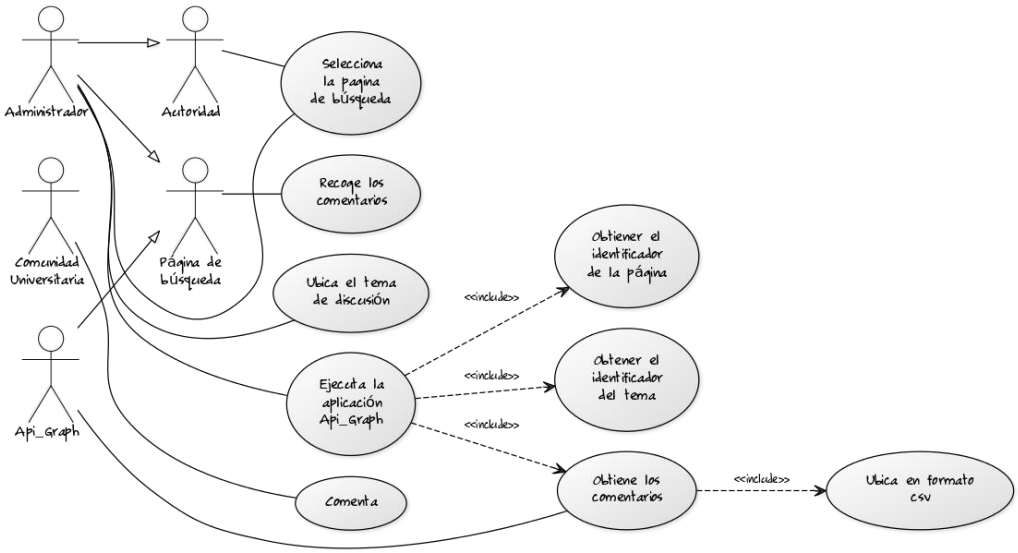


Figura 21: Caso de uso general
Elaborado por: Investigador

CASO PRÁCTICO

Objetivo

Aplicar la metodología de la propuesta utilizando como caso práctico los comentarios en el sitio oficial de Facebook de la UTA sobre el servicio académico "MATRICÚLATE YA"

Alcance

Opiniones vertidas en el Facebook de la UTA, sobre el servicio académico "MATRICÚLATE YA"

Procesos

1. **Selección de página y obtención de comentarios**
2. **Filtrado de las opiniones**
3. **Análisis mediante PMI**
4. **Resultados de la Clasificación**

Desarrollo

1. Selección de página y obtención de comentarios

El administrador del sistema se convierte en el principal actor del proceso, es quien junto a las autoridades universitarias selecciona el campo de acción (temática a proponer) en facebook con el fin de que la comunidad comente. Estos comentarios son la base fundamental para el analisis de opiniones y la clasificación de los servicios.

a) Selección de la temática, Ingresar al sitio oficial de la UTA y seleccionar la temática

Figura 22: Selección de la temática caso de estudio
Elaborado por: Investigador

Se propone analizar el servicio "MATRICÚLATE YA" que ofrece la UTA a través de la Dirección de Tecnología.

b) **Obtención de comentarios,** Para ello se utiliza el Api_Graph de Facebook https://developers.facebook.com/tools/explorer/ e iniciar sesión con la cuenta de facebook

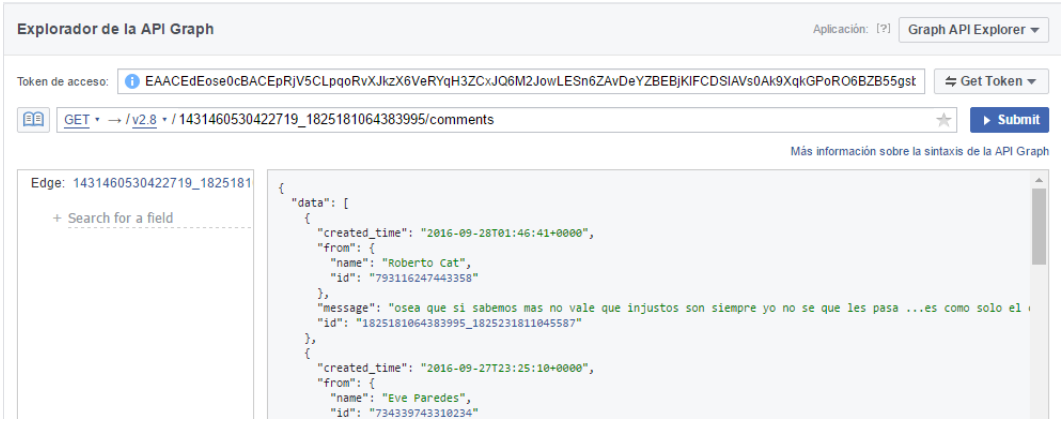
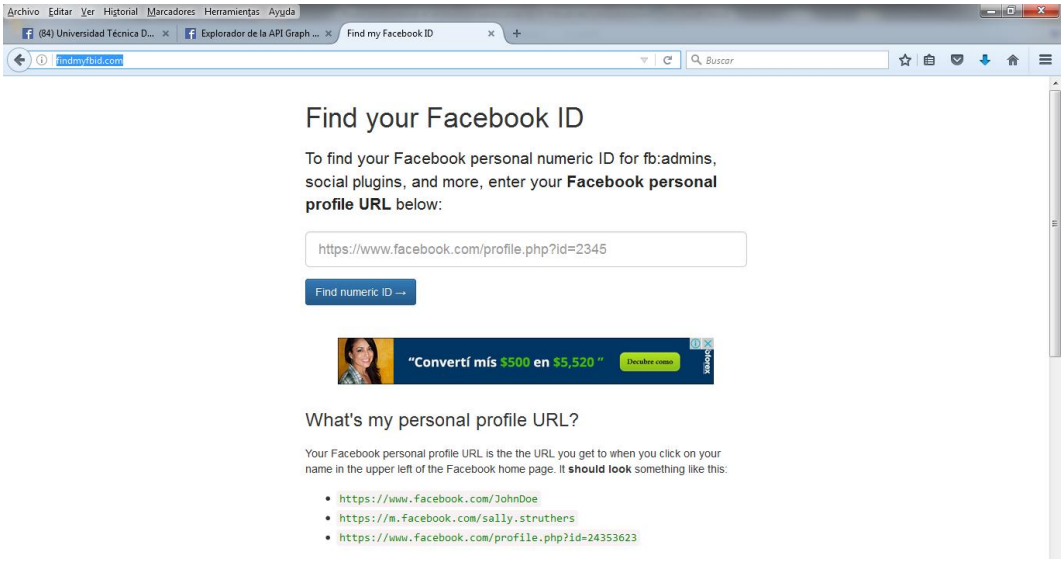


Figura 23: Obtención de comentarios Caso Práctico
Elaborado por: Investigador

La numeración: 1431460530422719, se corresponde con el identificador del tema. Este numero se obtiene ingresando a http://findmyfbid.com/



Y en el lugar que solicita dirección se ingresa la dirección web de facebook de la UTA :https://www.facebook.com/UniversidadTecnicadeAmbatoOficial/

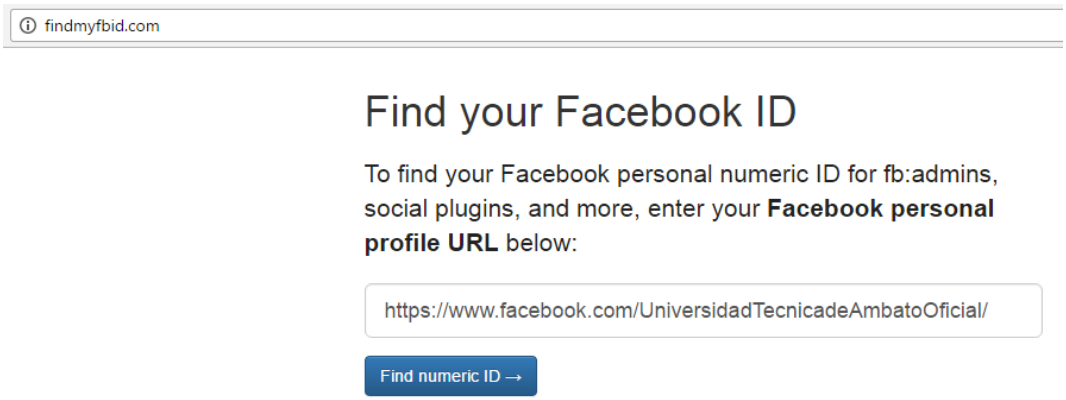


Figura 24: ID sitio oficial de la UTA
Elaborado por: Investigador

Luego al presionar "Find numeric ID" se obtiene como resultado 1431460530422719 (para el caso de la UTA)

Luego se busca una opinión en el sitio facebook, se presiona sobre la fecha y se copia el número que se genera al final, después de _id; ejemplo _id=1892542044314563

Figura 25: Identificador de comentarios
Elaborado por: Investigador

Al ejecutar todo el proceso, se obtiene los comentarios en JSON:

- Pulsar en "Obtener" y "Obtener token de acceso de usuario"
- Por último clic en enviar y se genera el documento con las opiniones

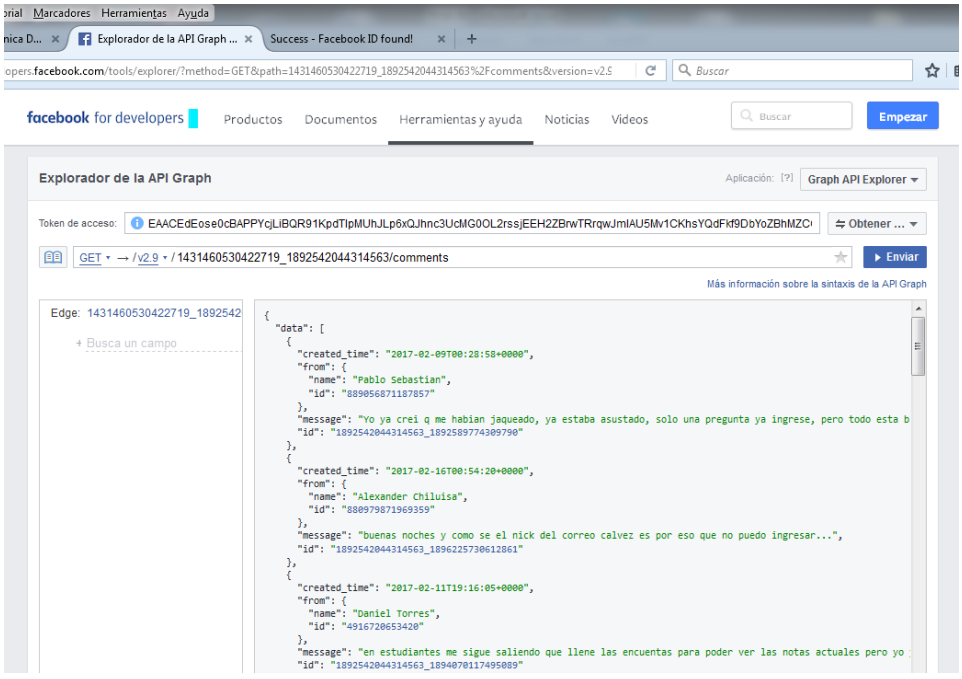


Figura 26: Obtención de todos los comentarios
Elaborado por: Investigador

Exportación a CSV

- Se utiliza la aplicación Konklone: https://konklone.io/json/
- Copiar y Pegar los comentarios obtenidos en API de facebook (Figura 26)

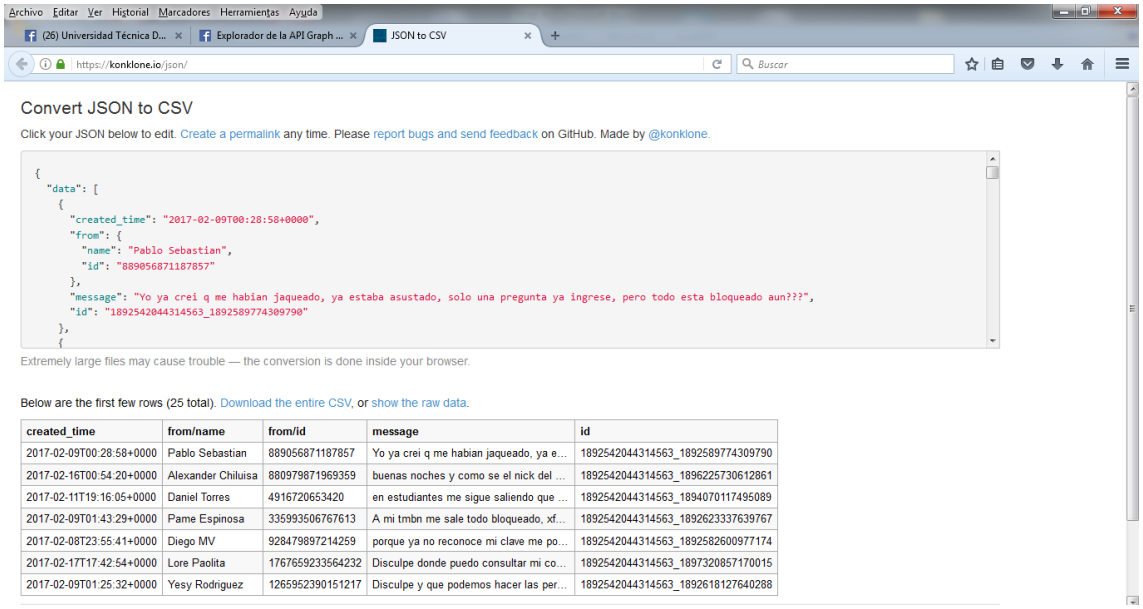


Figura 27: Aplicación Konklone caso de estudio
Elaborado por: Investigador

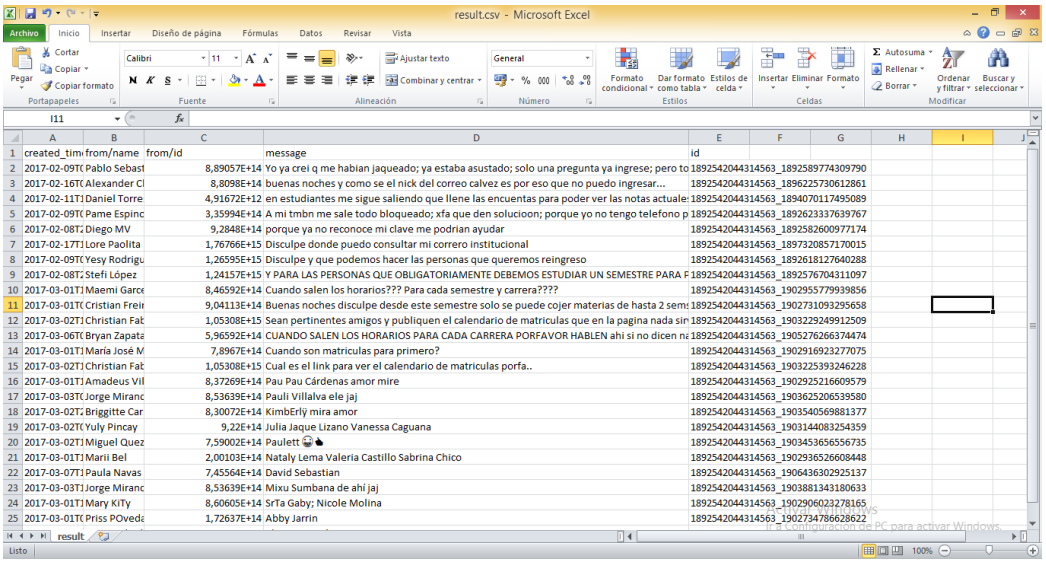
- Presionar en “Download the entire CSV” para obtener el archivo csv

Las opiniones del CSV deben ser filtradas por el Administrador de la aplicación antes de proceder a ejecutar la polaridad y el analisis de sentimientos previo a la clasificación por clusters de los servicios.

2. Filtrado de las opiniones

No todas las opiniones que se vierten para el tema son válidas, esto implica que el administrador ayudado por un experto en el tema, seleccione aquellas opiniones que van a formar parte del conjunto que serán analizadas mediante PMI y el análisis de sentimientos, para lo cual se debe realizar lo siguiente:

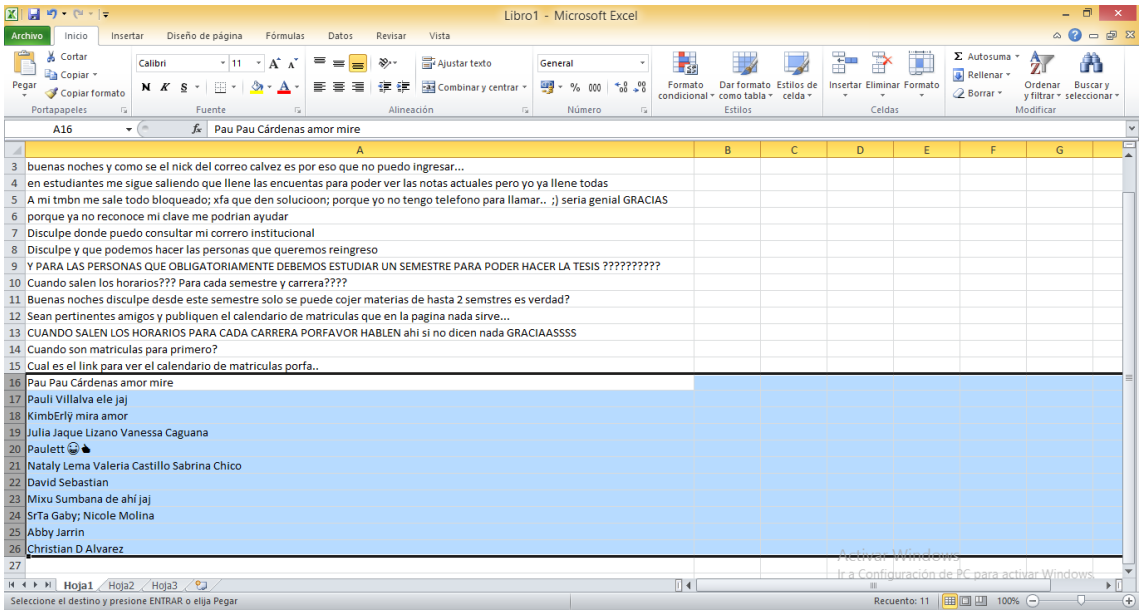
a) Abrir el archivo csv obtenido con los comentarios



created_tim	from/name	from/id	message	id
2017-02-09T(	Pablo Sebast	8,89057E+14	Yo ya creí q me habían jaqueado; ya estaba asustado; solo una pregunta ya ingrese; pero to	1892542044314563_1892589774309790
2017-02-16T(	Alexander Cl	8,8098E+14	buenas noches y como se el nick del correo calvez es por eso que no puedo ingresar...	1892542044314563_1896225730612861
2017-02-11T(	Daniel Torre	4,91672E+12	en estudiantes me sigue saliendo que llene las encuentas para poder ver las notas actuale	1892542044314563_1894070117495089
2017-02-09T(	Pame Espinc	3,35994E+14	A mi tmbn me sale todo bloqueado; xfa que den solucioon; porque yo no tengo telefono p	1892542044314563_1892623337639767
2017-02-08T(	Diego MV	9,2848E+14	porque ya no reconoce mi clave me podrian ayudar	1892542044314563_1892582600977174
2017-02-17T(	Lore Paolita	1,76766E+15	Disculpe donde puedo consultar mi correro institucional	1892542044314563_1897320857170015
2017-02-09T(	Yesy Rodrigu	1,26595E+15	Disculpe y que podemos hacer las personas que queremos reingreso	1892542044314563_1892618127640288
2017-02-08T(	Stefi López	1,24157E+15	Y PARA LAS PERSONAS QUE OBLIGATORIAMENTE DEBEMOS ESTUDIAR UN SEMESTRE PARA F	1892542044314563_1892576704311097
2017-03-01T(	Maemi Garce	8,46592E+14	Cuando salen los horarios??? Para cada semestre y carrera????	1892542044314563_1902955779939856
2017-03-01T(	Cristian Frei	9,04113E+14	Buenas noches disculpe desde este semestre solo se puede cojer materias de hasta 2 sem	1892542044314563_1902731093295658
2017-03-02T(	Christian Fat	1,05308E+15	Sean pertinentes amigos y publiquen el calendario de matriculas que en la pagina nada sir	1892542044314563_1903229249912509
2017-03-06T(	Bryan Zapata	5,96592E+14	CUANDO SALEN LOS HORARIOS PARA CADA CARRERA PORFAVOR HABLEN ahi si no dicen na	1892542044314563_1905276266374474
2017-03-01T(	María José M	7,8967E+14	Cuando son matriculas para primero?	1892542044314563_1902916923277075
2017-03-02T(	Christian Fat	1,05308E+15	Cual es el link para ver el calendario de matriculas porfa..	1892542044314563_1903225393246228
2017-03-01T(	Amadeus Vil	8,37269E+14	Pau Pau Cárdenas amor mire	1892542044314563_1902925216609579
2017-03-03T(	Jorge Miranc	8,53639E+14	Pauli Villalva ele jaj	1892542044314563_1903625206539580
2017-03-02T(	Briggitte Car	8,30072E+14	KimbErly mira amor	1892542044314563_1903540569881377
2017-03-02T(	Yuly Pincay	9,22E+14	Julia Jaque Lizano Vanessa Caguana	1892542044314563_1903144083254359
2017-03-02T(	Miguel Quez	7,59002E+14	Paulett	1892542044314563_1903453656556735
2017-03-01T(	Marii Bel	2,00103E+14	Nataly Lema Valeria Castillo Sabrina Chico	1892542044314563_1902936526608448
2017-03-07T(	Paula Navas	7,45564E+14	David Sebastian	1892542044314563_1906436302925137
2017-03-03T(	Jorge Miranc	8,53639E+14	Mixu Sumbana de ahí jaj	1892542044314563_1903881343180633
2017-03-01T(	Mary KiTy	8,60605E+14	SrTa Gaby; Nicole Molina	1892542044314563_1902906023278165
2017-03-01T(	Priss POveda	1,72637E+14	Abby Jarrin	1892542044314563_1902734786628622

Figura 28: Nuevos comentarios
Elaborado por: Investigador

1. Filtrar manualmente las opiniones (solo nos sirve la columna message)



buenas noches y como se el nick del correo calvez es por eso que no puedo ingresar...
en estudiantes me sigue saliendo que llene las encuentas para poder ver las notas actuales pero yo ya llene todas
A mi tmbn me sale todo bloqueado; xfa que den solucioon; porque yo no tengo telefono para llamar.. ;) seria genial GRACIAS
porque ya no reconoce mi clave me podrian ayudar
Disculpe donde puedo consultar mi correro institucional
Disculpe y que podemos hacer las personas que queremos reingreso
Y PARA LAS PERSONAS QUE OBLIGATORIAMENTE DEBEMOS ESTUDIAR UN SEMESTRE PARA PODER HACER LA TESIS ??????????
Cuando salen los horarios??? Para cada semestre y carrera????
Buenas noches disculpe desde este semestre solo se puede cojer materias de hasta 2 semstres es verdad?
Sean pertinentes amigos y publiquen el calendario de matriculas que en la pagina nada sirve...
CUANDO SALEN LOS HORARIOS PARA CADA CARRERA PORFAVOR HABLEN ahi si no dicen nada GRACIAASSSS
Cuando son matriculas para primero?
Cual es el link para ver el calendario de matriculas porfa..
Pau Pau Cárdenas amor mire
Pauli Villalva ele jaj
KimbErly mira amor
Julia Jaque Lizano Vanessa Caguana
Paulett
Nataly Lema Valeria Castillo Sabrina Chico
David Sebastian
Mixu Sumbana de ahí jaj
SrTa Gaby; Nicole Molina
Abby Jarrin
Christian D Alvarez

Figura 29: Filtrado manual de comentarios
Elaborado por: Investigador

Como se observa en la Figura 29, es necesario eliminar comentarios que se consideran no válidos como los nombres de las personas. Estos nuevos comentarios (9), van a ser ubicados

en cada fichero y además sumados a los 71 que se toma como entrenamiento. Se podrán ubicar mayor cantidad de comentarios en el futuro según vaya llenándose la página de los mismos. Además y dado que trabajamos con un diccionario en ingles estos serán convertidos a ese idioma, utilizando para ello cualquier traductor:

I alreadythoughttheyhad hit me; He wasalreadyscared; Justonequestion and enter; Buteverythingisstillblocked ???Goodnight and as the mail nick of calvez iswhy I can notenter ... In students I stillleave to fillthem to seethecurrent notes but I alreadyfillall To me tmbn me sale allblocked; Xfagivingsolution; Because I do nothave a phone to call..;) itwould be great THANK YOU Becauseyou no longerrecognizemypassword, theycouldhelp me. Excuse me where I can seemyinstitutionalcorero Excuse me and what can peopledoingthatwant to re-enter AND FOR PEOPLE WHO MUST OBLIGATORY TO STUDY A SEMESTER TO BE ABLE TO DO THE THESIS ??????????When do theschedulesleave ???Foreachsemester and career ????Goodeveningfromthissemesteryou can onlytakethesubjects of up to 2 semesterisit true? Be friends and publishtheregistration calendar thatonthe page nothingserves ... WHEN THE SCHEDULES LEAVE FOR EACH CAREER PLEASE SPEAK THERE ifthey do notsayanything GRACIAASSSS When are theyenrollmentforfirst? Whatisthe link to seetheregistrationschedule porfa ..

Figura 30: Comentarios en ingles
Elaborado por: Investigador

El diccionario en ingles proviene de la librería TM. Este diccionario se encuentra probado para la terminología inglesa. Se ubica de esta manera en el código fuente del archivo PMI:

```
library(tm)

#Load thespeechesfrom a directory and applysome simple cleanuptransformations
```

Figura 31: Diccionario en ingles
Elaborado por: Investigador

Con los comentarios que se reflejan en la figura anterior vamos a tener 80 elementos para trabajar la nueva clasificación.

3. Análisis mediante PMI

a) Utilización del programa R

- En R se abre el archivo PMI.R (archivo elaborado y adjunto a la propuesta que debe ser cambiado el número de etiquetas en el código fuente a 80, pues anteriormente para entrenar se trabajó solo con 71.)
- Se cambia el path de donde están los comentarios en csv y en "Editar" se selecciona "Ejecutar Todo".

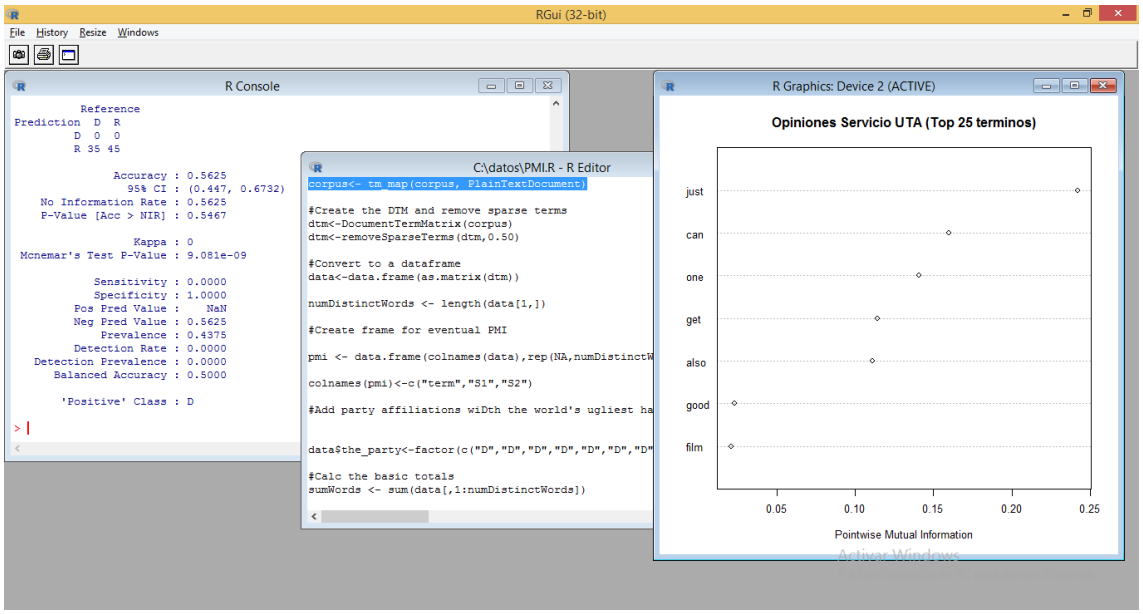


Figura 32: Resultados primarios caso de estudio
Elaborado por: Investigador

4. Resultados de la Clasificación

Existen mayor cantidad de comentarios positivos que negativos sobre el servicio "MATRICÚLATE YA", tomando en cuenta que se trabaja con datos de entrenamiento, con una precisión actual de 0.56 sobre falsos negativos y una especificidad sobre los falsos positivos del 100%. En la gráfica se muestran los resultados de los términos que más son utilizados en los comentarios y que pueden servir a futuro para formar un nuevo diccionario de palabras en español:

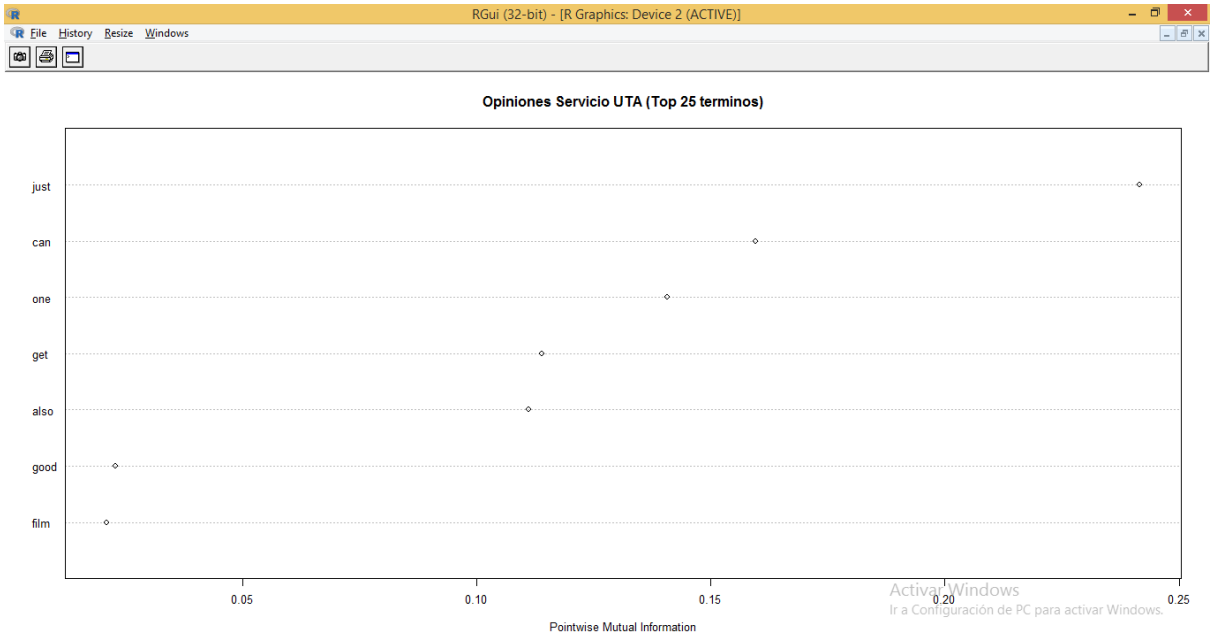


Gráfico 7: Términos más utilizados
Elaborado por: Investigador

Como se observa la palabra **just** es la más utilizada para cada uno de los comentarios.

CONCLUSIONES

- La libertad como se obtienen los comentarios marca la diferencia con las entrevistas. En este caso al proceder los comentarios de una red social, es libertad de cada uno de los autores ubicar su favoritismo o negatividad hacia un servicio, esto hace que exista la mayor claridad sobre la calidad del mismo.
- El filtrado de la información es necesario para construir el fichero de entrenamiento y prueba, puesto que, en los comentarios al ser libres tambien pueden recibir cualquier frase que no necesariamente involucre un concepto de calidad.
- Se puede realizar agrupamiento de opiniones. En este caso se logra un agrupamiento de positivos y negativos y se logra decidir cual de los dos tiene mas peso en el conjunto de opiniones de los usuarios, tomando en cuenta que; en un mismo párrafo pueden existir opiniones positivas y negativas.
- La precisión de un sistema en el cual es libre las opiniones, no puede llegar a un 100%. El comportamiento humano es impredecible por ende también sus comentarios y opiniones la cual depende estrictamente de como es atendido y de la calidad del servicio.
- En un sistema en donde las observaciones-polaridad de las opiniones son bivalentes (positivas o negativas) la asignacion de las mismas dependen directamente de la opinion de un experto, y es necesario enseñar esa precisión de experto al sistema con el fin de que pueda predecir con claridad la polaridad de un texto y por ende la calidad del servicio.
- La precisión obtenida mayor al 65% en la mayoría de los indicadores se considera como válida para el sistema puesto que parte de opiniones libres de contexto sobre la calidad del servicio.

- La clasificación por medio del algoritmo Kmeans permitió identificar a estos grupos, lo que implica que esta investigación puede ser sometida para el análisis de otros servicios académicos dentro de la universidad. El criterio estadístico

permitió corroborar lo obtenido por el clasificador por lo que se considera a las variables (atributos) como adecuados para medir la satisfacción de los servicios según la calidad.

RECOMENDACIONES

- Se debe trabajar con espacios libres de contexto y en este caso investigar con mayor detalle sobre la libertad de comentarios en las redes sociales, puesto que de esto se alimenta el big data en este tipo de entornos.
- Se debe investigar a mayor detalle el filtrado de la información, quiza estructurando la información con el fin de que a partir de esta el big data tome forma, y los parametros obtenidos puedan ser filtrados por métodos de minería de datos o texto.
- Se puede mejorar el agrupamiento de las opiniones no necesariamente para dos grupos (positivo y negativo) sino incluir el agrupamiento de like en big data en redes sociales (me gusta, no me gusta, asombrado y mas).
- Se puede mejorar la precisión del sistema con mayor cantidad de opiniones, las cuales se generan en el entorno del big data red social. Esto montando grandes campañas sobre la calidad de los servicios.
- El incluir mayor cantidad de observadores mejora el resultado del indice de kappa. De hecho para estos sistemas se puede considerar tener al menos 4 observadores expertos que vayan etiquetando la polaridad de los textos con el fin de mejorar la precisión del sistema.
- Si bien el porcentaje de precisión obtenido es mayor al 65%, se puede incrementar la misma mediante técnicas de procesado de información, lo que involucraría nuevos estudios que resulten a partir de la presente tesis de maestría**.**

BIBLIOGRAFIA

Adeline, A., Grouin, C., Pierre, Z., & Falissard, B. (2015). Text mining applications in pshchiatry: a systematic literature review. *Phsychiatric research*(DOI: 10.1002/mpr.1481).

Barranco, R. (18 de 06 de 2012).

https://www.ibm.com/developerworks/ssa/local/im/que-es-big-data/.

Bar-Yam, Y. (2016). From big data to important information. *Complexity*.

Blackburn, D., & Reuber, M. (2015). Using conversation analysis to help diagnose dementia. *Aging, Dementia, Cognitive, and Behavioral Neurology ePosters*(14), 84.

Cobo, A. R., & Martínez, M. (2009). Descubrimiento de conocimiento en repositorios documentales mediante tecnicas de Mineria de Texto y Swarm Intelligence. *Revista Electronica de Comunicaciones y Trabajos de ASEPUMA*, 105-124.

Corso, C. L. (2009). Aplicación de algoritmos de clasificación supervisada usando. *congresos_labsis*, 1-11.

David Vilares, Miguel A. Alonso. (s.f.). Clasificacion de polaridad en textos con opiniones en español. *Academia*.

Dubiau, L. (2013). Obtenido de http://materias.fi.uba.ar/7500/Dubiau.pdf

Hangya, V. B. (2013). *Sentiment Detection on Twitter Messages.*Obtenido de http://www.cs.york.ac.uk/semeval-2013/accepted/102_Paper.pdf

LaValle, S. M. (Lunes de Octubre de 2010). *Analytics: el uso de big data en el mundo real - IBM*. Obtenido de www-05.ibm.com/services/es/gbs/.../pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf: http://www-05.ibm.com/services/es/gbs/consulting/pdf/El_uso_de_Big_Data_en_el_mundo_real.pdf

Levallois, C. (2013). *Sentiment Analysis for Tweets based on Lexicons an Heuristics.* Obtenido de http://www.cs.york.ac.uk/semeval-2013/accepted/27_Paper.pdf

Liu, B. (2012). *Sentiment Analysis and Opinion Mining.*Morgan & Claypool.

Malacara, N. (6 de octubre de 2014). Obtenido de http://www.informabtl.com/5-caracteristicas-del-big-data/

Morate, D. G. (2000). *Manual de Weka.*

Morita, K., Fuketa, M., Aoe, J.-i., & Yasuda, K. (2015). Improved dialogue communication systems for individualas with dementia. *Computer Applications in Technology*, *52*(2-3), 127- 134.

Mulholland, M., & Quinn, J. (2013). Suicidal Tendencies: The Automatic Classification of Suicidal and Non-Suicidal Lyricists using NLP. *IJCNLP*, 680-684.

Nguyen, T., Hoang, H., Van, D., Van, T., & Duy, A. (2016). An Improvement of KMeans Algorithm Using Wavelet Technique to Increase Speed of Clustering Remote Sensing Images. *International Journal of Computer and Electrical Enginnering*, 177-184.

Onofri, E., Mercuri, M., Archer, T., Max Rap, R., Massoni, F., & Ricci, S. (2015). Effect of cognitive fluctuation of handwriting in alzheimer's patient: a case study. *Acta Medica Mediterranea, 31*(751).

Pang, B., L. Lee, y Vaithyanathan. (2002). Sentiment classification usingmachine learning techniques.

Poria, S., Gelbukh, A., Cambria, E., Hussain, A., & Guang-Bin, H. (2014). EmoSenticSpace: A novel framework for affective common-sense. *Elsevier*, 108-123.

Santamaría P., R., & Mejías A., A. (2013). Análisis de la calidad de los servicios académicos: estudio de caso en Universidad. *Revistas Científicas de América Latina y el Caribe, España y Portugal* , 67-74.

Saralegi, X., & San Vicente, I. (2013). Elhuyar at TASS 2013. *XXIX Congreso de la Sociedad Española de Procesamiento de lenguaje natural* (págs. 143-150). Madrid: SEPLN.

Scull, R., Thorup, J., & Howell, S. (2016). Review 10 trend impacting distance and continuing education. *Recruiting & Retaining*, 1-5.

Thelwall, M., Buckley, K., & Paltoglou, G. (2012). *Sentiment strength detection for the social web. Journal of the American Society for Information Science and Technology.*

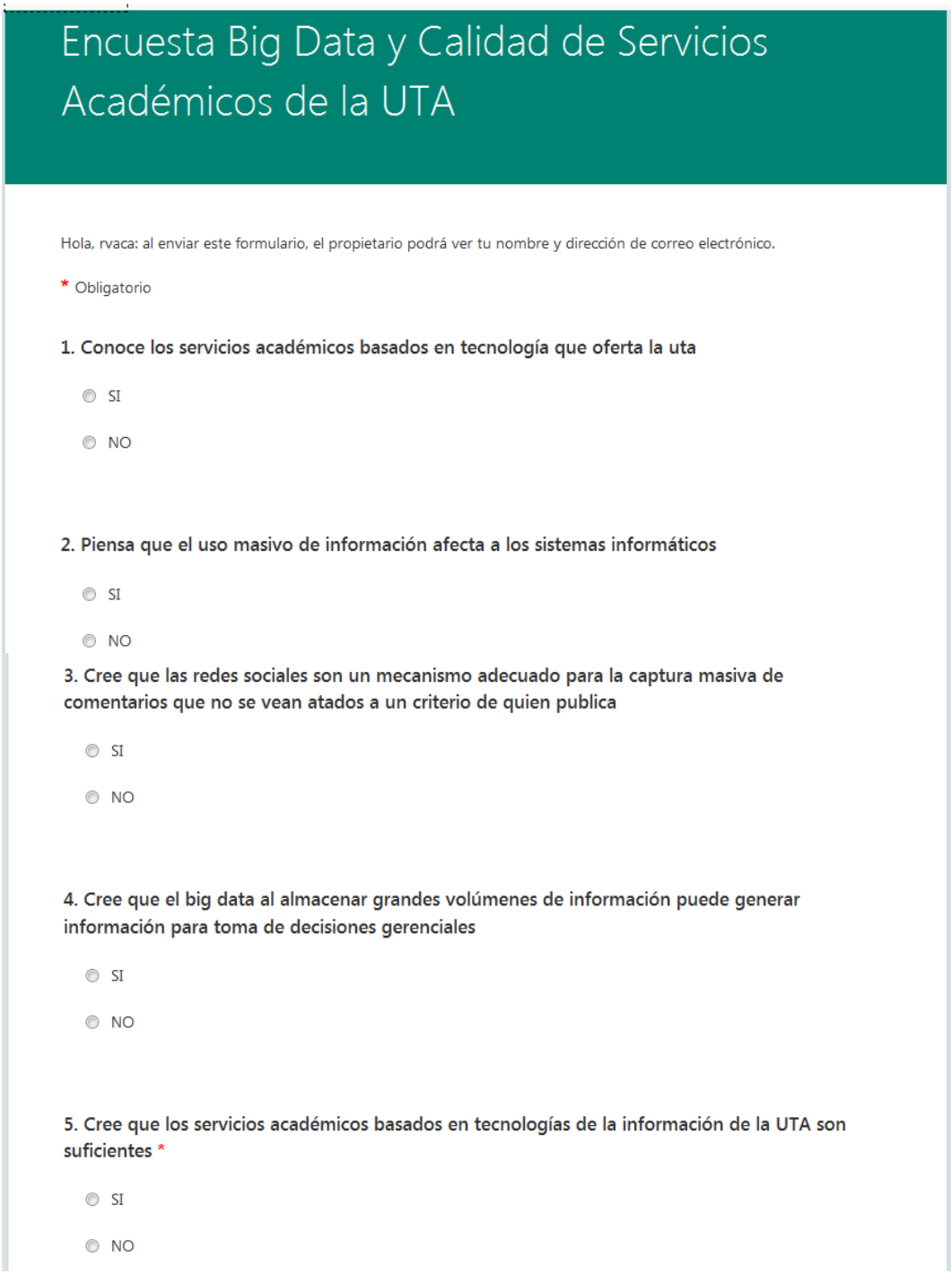
UMH. (s.f.). Obtenido de http://isa.umh.es/asignaturas/crss/turorialWEKA.pdf

Uriarte, R. D. (2003). *Introducción al uso y programación del sistema estadístico R.* Obtenido de https://cran.r-project.org/doc/contrib/curso-R.Diaz-Uriarte.pdf

Zhang, Y., Zhang, G., Chen, H., Porter, A., Zhu, D., & Lu, J. (2016). Topic analysis and forecasting for science, technology and innovation: Methodology with a case study focusing on big data research. *Technological Forecasting and Social Change*, 179-191.

ANEXOS

ANEXO 1.



Encuesta Big Data y Calidad de Servicios Académicos de la UTA

Hola, rvaca: al enviar este formulario, el propietario podrá ver tu nombre y dirección de correo electrónico.

* Obligatorio

1. Conoce los servicios académicos basados en tecnología que oferta la uta

- SI
- NO

2. Piensa que el uso masivo de información afecta a los sistemas informáticos

- SI
- NO

3. Cree que las redes sociales son un mecanismo adecuado para la captura masiva de comentarios que no se vean atados a un criterio de quien publica

- SI
- NO

4. Cree que el big data al almacenar grandes volúmenes de información puede generar información para toma de decisiones gerenciales

- SI
- NO

5. Cree que los servicios académicos basados en tecnologías de la información de la UTA son suficientes *

- SI
- NO

6. Posee usted una cuenta en la red social facebook

- SI
- NO

Enviar