

Información Importante

La Universidad Santo Tomás, informa que el(los) autor(es) ha(n) autorizado a usuarios internos y externos de la institución a consultar el contenido de este documento a través del catálogo en línea, página web y Repositorio Institucional del CRAI-USTA, así como en las redes sociales y demás sitios web de información del país y del exterior con las cuales tenga convenio la Universidad.

Se permite la consulta a los usuarios interesados en el contenido de este documento, para todos los usos que tengan **finalidad académica**, siempre y cuando mediante la correspondiente cita bibliográfica se le dé crédito al trabajo de grado y a su autor, nunca para usos comerciales.

De conformidad con lo establecido en el Artículo 30 de la Ley 23 de 1982 y el artículo 11 de la Decisión Andina 351 de 1993, la Universidad Santo Tomás informa que “los derechos morales sobre documento son propiedad de los autores, los cuales son irrenunciables, imprescriptibles, inembargables e inalienables.”

Centro de Recursos para el Aprendizaje y la Investigación, CRAI-USTA

Universidad Santo Tomás, Bucaramanga

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

**Patrones de comportamiento de los usuarios de la empresa Adonai Employment utilizando
minería de datos**

Jéssica Dayana Rojas Gómez

Trabajo de grado para optar el título de Ingeniera Industrial

Director

Javier Hernández Cáceres

Magister en Educación

Universidad Santo Tomás, Bucaramanga

División de Ingenierías y Arquitectura

Ingeniería Industrial

2020

Contenido

	Pág.
1. Definición del Problema.....	11
1.1 Planteamiento del Problema	11
1.2 Formulación del Problema	13
2. Justificación.....	13
3. Objetivos	15
3.1 Objetivo General.....	15
3.2 Objetivos Específicos.....	15
4. Marco Referencial.....	16
4.1 Marco Teórico	16
4.2 Marco Conceptual.....	17
4.2.1 Log	17
4.2.2 Log de Servidores	18
4.2.3 Log de Acceso	18
4.2.4 Dirección IP	19
4.2.5 Analog	19
4.2.6 Weka	19
4.2.7 Webmining	20

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

4.2.8	Minería de Datos (Data Mining).....	21
4.3	Estado del Arte	21
4.3.1	A Nivel Internacional.....	21
4.3.2	A Nivel Nacional	23
4.3.3	A Nivel Regional	26
5	Metodología.....	28
5.1	Tipo de Investigación.....	29
5.2	Diseño de la Investigación	30
5.3	Método	32
6.	Aplicación de las Herramientas de Web Mining para la Interpretación de la Información de Log de Acceso	33
7.	Resultados.....	35
7.1	Reportes e Informes con Analog	35
6.1.1	Informe de Organización.....	38
6.1.2	Informe de Sistemas Operativos.....	39
6.1.3	Informe de Códigos de Estado.....	40
6.1.4	Informe de Tamaño de Archivos	40
6.1.5	Informe de Peticiones.....	41
8.	Limpieza del Log	43
9.	Parametrización del Log.....	44

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

10.	Análisis en el Weka	49
11.	Reglas de Asociación Generadas por el Algoritmo A Priori en Weka.....	50
11.1	Análisis de las Reglas de Asociación.....	51
12.	Conclusiones	53
13.	Recomendaciones	56
	Referencias.....	58

Lista de Tablas

	Pág.
Tabla 1. <i>Método de Investigación</i>	32
Tabla 2. <i>Listado de los sistemas operados ordenados por número de peticiones por páginas</i> ...	39
Tabla 3. <i>Listado de archivos que representan por lo menos 20 peticiones</i>	42
Tabla 4. <i>IP Usuarios Log</i>	44
Tabla 5. <i>Fecha de Acceso Log</i>	44
Tabla 6. <i>Bytes Transferidos</i>	45
Tabla 7. <i>Navegador Utilizado Log</i>	45
Tabla 8. <i>URL de Búsqueda Log</i>	45

Lista de Figuras

	Pág.
<i>Figura 1.</i> Log Sucio Adonai Employment.....	34
<i>Figura 2.</i> Log Copiado en Carpeta Analog 6.0	39
<i>Figura 3.</i> Analog.cfg.....	36
<i>Figura 4.</i> Resumen General Analog Adonai Employment	36
<i>Figura 5.</i> Resumen Diario Adonai Employment.....	37
<i>Figura 6.</i> Resumen Horario Adonai Employment	38
<i>Figura 7.</i> Informe de la Organización Adonai Employment	38
<i>Figura 8.</i> Informe de Códigos de Estado	40
<i>Figura 9.</i> Informe de Tamaños de Archivos	45
<i>Figura 10.</i> Informe de Peticiones	45
<i>Figura 11.</i> Log Limpio Adonai Employment	45
<i>Figura 12.</i> Log Parametrizado Adonai Employment	45
<i>Figura 13.</i> Documento Extensión .Arff.....	45
<i>Figura 14.</i> Gráfica Columna IP.....	49
<i>Figura 15.</i> Gráficas de los 9 Atributos	50

Resumen

En esta investigación, se realiza el manejo, depuración y aplicación de Webmining y Datamining a partir de un log generado por un servidor web perteneciente a Adonai Employment, agencia de empleos cristiana, sin ánimo de lucro y libre de impuestos que presta sus servicios en el área de Milwaukee a personas que poseen discapacidades físicas. Este registro de eventos hace referencia al análisis de los datos obtenidos en el período del 06 al 27 de febrero de 2020, por lo cual, en primer lugar, se accede a la base de datos de la empresa a través del log, posteriormente se analizan los datos con el software Analog y Weka y, finalmente se convierten en información útil para la empresa, la cual ayuda a conocer los patrones de comportamientos en cuanto acceso a la web y servicios más solicitados por el usuario.

Para el desarrollo del proyecto se utilizan reglas de asociación (herramientas de la minería de datos) que permite cumplir con el objetivo de caracterizar a los clientes de Adonai Employment. A su vez, la empresa puede conocer factores importantes como los días de la semana en los que su página web tuvo mayor flujo y la hora del día más concurrida, entre otras variables expuestas a lo largo del documento.

Palabras Clave: Webmining, Datamining, log, Analog, caracterizar.

Abstrac

In this research, the management, purification and application of Webmining and Datamining is carried out from a log generated by a web server belonging to Adonai Employment, Christian employment agency, nonprofit, tax-free services in the Milwaukee area to people with physical disabilities. This event register refers to the analysis of the data obtained in the period from 06 to 27 February 2020, so that, first of all, the company's database is accessed through the log, Later the data are analyzed with the Analog software and finally become useful information to the company, which helps to know the behavioral patterns regarding access to the web and services most requested by the user.

For the development of the project, partnership rules (data mining tools) are used to meet the objective of characterizing Adonai Employment's clients. At the same time, the company can know important factors such as the days of the week in which its website had the highest flow and the busiest time of day, among other variables exposed throughout the document.

Key Words: Webmining, Datamining, log, Analog, characterize.

Introducción

La sociedad actual goza de un privilegio que anteriormente no se poseía: la globalización, la cual juega un papel importante en el manejo de las finanzas mundiales, ya que permite extender el comercio de una ciudad que antes se limitaba a realizar sus ventas de manera local a hacerlo de manera global, mediante la implementación de herramientas como el internet. Este a su vez, ofrece a las personas la posibilidad de acceder a diferentes lugares del mundo sin necesidad de salir de su lugar de origen, brindándoles con ello la facilidad de comercializar económicamente sus mercancías.

Gracias a esta herramienta se ha hecho posible ofrecer gran variedad de productos que satisfacen las necesidades de las personas, puesto que, en internet con las redes sociales, los blogs y los diferentes espacios de opinión y expresión es posible obtener datos que entregan de una manera u otra información sobre los gustos y necesidades de las personas; el problema, sin embargo, se encuentra en la interpretación de los datos generados.

Es por ello que surgen diferentes técnicas que al ser implementadas permiten analizar los datos generados por las acciones de las personas, los cuales, de manera implícita contienen la información necesaria para hacer un análisis detallado tanto de las necesidades del usuario como de la evolución y el manejo que se le hace a la empresa, para finalmente corregir los errores que pueden llevar al fracaso la misma.

Para la realización del presente trabajo se hace necesaria la implementación de dos técnicas importantes en la extracción de la información, las cuales son conocidas como data mining y web mining. Estas técnicas son procesos fundamentales para todo tipo de organización ya que permiten

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

adquirir información referente a las estadísticas de sus páginas web, tales como el número de visitas diarias, el día a la semana con mayor actividad, información de la persona u organización que realiza la consulta, el sistema operativo que maneja, así como los bytes transferidos.

En el presente proyecto, la minería de datos facilita la obtención de información histórica de la empresa Adonai Employment, la cual es analizada para determinar los patrones de comportamientos de los clientes. Como se trata de una empresa de servicios sin ánimo de lucro, se obtienen datos de los días de la semana con mayor cantidad de visitas y horas del día más frecuentadas, asimismo, se obtiene información estadística, incluso de los navegadores utilizados, sistemas operativos y de la forma en que se realiza la búsqueda y se dirigen a la página www.adonai-employment.org, ya que en algunos casos el acceso es inválido o se presentan problemas en el servidor. Todo esto permite que se examinen falencias presentes en la página web y finalmente se determinen estrategias para ofrecer una buena impresión y un servicio de calidad desde el momento en que ingresan a la página web, contribuyendo de igual manera al crecimiento de la empresa Adonai Employment.

1. Definición del Problema

1.1 Planteamiento del Problema

En los últimos años, ha incrementado la capacidad para generar y almacenar información en los servidores web, sin embargo, la capacidad de darle uso y procesar estos datos no ha ido a la par. Cada una de las solicitudes realizadas por un usuario queda registrada como una huella que los servidores automáticamente almacenan en una bitácora de accesos (log), por ende, es posible coleccionar gran cantidad de datos. No obstante, al tener tantos datos, mucha información se encuentra

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

oculta, por tal razón, para poder entender el sentido de todos estos datos y la relación entre ellos se deben utilizar aplicaciones de Minería de Datos, con el objetivo de filtrar los datos de acuerdo a aquella información que se quiere analizar, lo cual permite que sea más sencillo concentrarse en la información relevante, interpretar los resultados obtenidos y percibir la situación actual de la empresa, para finalmente corregir los errores que pueden afectar el buen desempeño de la misma y tomar decisiones cuyo objetivo sea su beneficio.

Los servidores web registran las entradas en archivos log, donde almacenan información relacionada con I.P del servidor, I.P del cliente, fecha y tiempo de acceso, enlaces visitados, bytes transferidos, bytes recibidos, navegadores utilizados y sistemas operativos del cliente. En realidad, la información que se almacene en los log depende de la configuración que se haga al servidor web. Cabe resaltar la importancia que tiene para una empresa conocer esta información registrada con los archivos log, con el fin de precisar la calidad del servicio de su sitio web, en cuanto a eficiencia y eficacia.

Las páginas web son usadas en la actualidad por las compañías, principalmente grandes empresas, sin embargo, cada vez son más las Pymes colombianas (para el 2016 correspondían al 55%) que hacen uso de este servicio, lo que se puede traducir como una oportunidad de identificar en la base de datos, comportamientos de los clientes que no son muy evidentes, en algunos casos porque no se emplean herramientas de medición web, mientras que en otros se debe a que no se hace una buena lectura de estas herramientas o aplicaciones.

La carencia de herramientas de medición web que permitan obtener y analizar de manera clara y precisa los comportamientos de los usuarios en compañías en las que sus operaciones tienen un sustento digital como lo son las páginas web, trae consigo el desaprovechamiento de información que revela factores fundamentales para la continua mejora, evolución y actualización

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

de sus servidores, además desperdician la posibilidad de conocer detalles relevantes de sus clientes que le permitan una clara interpretación de diferentes aspectos, los cuales pueden facilitar el acercamiento con los clientes de la compañía, dejándola en desventaja frente a la competencia.

De esa manera, la empresa Adonai Employment podrá tener conocimiento de todo lo que ocurre detrás de su página web a través de las técnicas y herramientas de minería de datos, obteniendo información clave que oriente sobre el diseño de las páginas, tipo de información que debe ser publicada, horarios en los que es preciso no recargar el servidor, detección de enlaces rotos o archivos no disponibles, la frecuencia y periodicidad para actualizaciones y estrategias. Además, le va a permitir encontrar, atraer y retener los clientes conociendo particularmente la tendencia de sus comportamientos, para identificar y anticipar sus necesidades lo que también contribuye a una mejoría en la relación con los mismos, permitiendo que la información extraída constituya un instrumento muy útil para la toma de decisiones estratégicas que guíen al crecimiento y fortalecimiento de la compañía.

1.2 Formulación del Problema

¿Cuáles son los patrones de comportamiento de navegación en la web de los usuarios de la empresa Adonai Employment?

2. Justificación

El objetivo fundamental del presente proyecto, consiste en identificar las características de los usuarios que acceden a la página web de la empresa Adonai Employment, para así reconocer las tendencias presentes en los mismos, el objeto por el cual realizan dicha consulta, el autor de la

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

consulta, el tipo de empresa (en caso de que se trate de una); patrones pertinentes para ser analizados por Adonai Employment y poder aplicar estrategias orientadas al crecimiento como compañía, lo que beneficia a múltiples partes.

Como se mencionaba anteriormente, el estudio a realizar tiene como finalidad el beneficio de diversas partes, principalmente de las personas del área metropolitana de Milwaukee con cierta desventaja o discapacidad, de los empleadores y finalmente, la empresa en conjunto con el país, puesto que “atiende al interés general o beneficia a un número importante de personas, de forma tal que las mismas contribuyen a la realización de los fines del Estado, al adelantar actividades que el Estado no alcanza a desarrollar” [1].

Adonai Employment, plantea que es una empresa en crecimiento, debido a que, en el año 2012, tuvo una carga de casos 400% mayor que en su primer año. Es esta, una de las razones por la cual se realiza el estudio a través del uso de la minería de datos, ya que representará un gran beneficio para la empresa poder abarcar mayor cantidad de personas, las cuales tendrán la oportunidad que en muchas ocasiones no les es concebida por su discapacidad [2].

Con relación a las personas discapacitadas, luego de los estudios necesarios en el ámbito de web mining, se pretende incrementar el porcentaje de estas personas con empleo formal, ya que, en muchos casos, difícilmente llegan a tener estabilidad laboral o seguramente, no les conceden la oportunidad de un trabajo digno y Adonai Employment a través de un seguimiento puede contactar con mayor facilidad a quien cumpla con las características y necesite un apoyo por parte de la empresa. A su vez, los empleadores también se benefician, puesto que diversifican la fuerza laboral de su organización y a la par, “les dan el beneficio de reducir los costos de la nómina proporcionando candidatos calificados para el Crédito Fiscal de Oportunidades de Trabajo (WOTC)” [2].

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Asimismo, cabe resaltar, que Adonai Employment, tiene principalmente la finalidad de velar por el bien común, a través de la implementación de técnicas y alianzas con más de 250 empleadores, donde han sido más de 600 personas las que han podido superar la barrera, tener un empleo y mejorar su calidad de vida. Para la empresa es de gran importancia conocer los patrones de comportamiento de los usuarios porque son parte del éxito de esta, por ende, hay que dar respuesta a sus necesidades e intereses, realizando modificaciones en cuanto al contenido y/o la estructura de su página web, de esta manera, puede resultar poco más sencillo para toda persona, tanto empleadores como solicitantes de empleo acceder a www.adonai-employment.org y ponerse en contacto o requerir sus servicios de apoyo en el momento que sea necesario.

3. Objetivos

3.1 Objetivo General

Caracterizar los patrones de comportamiento de los usuarios de la empresa Adonai Employment extraídos de un log del servidor por medio de técnicas de minería de datos.

3.2 Objetivos Específicos

- Describir los patrones de comportamiento de navegación de los usuarios de la empresa Adonai Employment utilizando el software Analog.
- Implementar una sábana de datos con el log del servidor de la empresa Adonai Employment por medio de un data warehouse para conocer las tendencias de navegación y características de los usuarios utilizando reglas de asociación.
- Validar los resultados de navegación del usuario generados por el software Analog con los resultados del software Weka.

4. Marco Referencial

4.1 Marco Teórico

En el año 2018, se llevó a cabo el proyecto “Establecimiento de patrones de comportamiento de los usuarios de la web American Motos S.A” por Cristian Leonardo Pineda Tarazona, estudiante de Ingeniería Industrial de la Universidad Santo Tomás, con el objetivo de identificar las tendencias de los clientes de esta empresa, la cual tiene disponibilidad en todo el país desde 1994. A través del estudio fue posible obtener datos tales como los días de la semana en los que los usuarios más frecuentaban la página web, el producto que los usuarios más visitaban, el horario en el que los clientes más compras realizaban, entre otros. Finalmente, se plantearon estrategias que la empresa debió adoptar para incrementar sus ganancias y generar mayor reconocimiento de la marca [3].

En la misma dirección, los resultados del estudio de Víctor Alfonso Sanabria, Ingeniero Industrial de la Universidad Industrial de Santander, para el año 2017, tras plantear la aplicación de técnicas de agrupamiento para el análisis estadístico de tendencias en twitter basado en el lenguaje de programación R. y la utilización de la técnica Análisis de Clúster, indican que regularmente se presentan situaciones tales como los errores ortográficos cometidos por las personas al escribir algún tweet, asimismo se pudo captar información acerca de los temas con mayor relevancia en la red social twitter y la gran cantidad de seguidores desde el punto de vista del mercado objetivo de la cuenta, finalmente Víctor pudo cumplir con su objetivo de “clasificar documentos en bases de datos del benchmarking” haciendo uso del twitter [4].

A su vez, también se aplicó la minería de datos en el ámbito de la salud en 2017 por Erika Rojas y Juan Sebastián Aguilar, con el propósito de optar al título de Ingeniero de Sistemas en este mismo año, por lo cual publican su trabajo de grado titulado “Minería de datos para el

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

descubrimiento de patrones en enfermedades respiratorias en Bogotá, Colombia”, para el desarrollo de su investigación utilizaron el algoritmo de clustering K-means a través de la herramienta RapidMiner, para poder identificar posibles comportamientos en enfermedades respiratorias en Bogotá. La minería de datos en el área de las ciencias de la salud, es utilizada para detectar y prevenir enfermedades, para analizar marcadores genéticos, prever la probabilidad de una respuesta satisfactoria a un tratamiento médico, así como para determinar factores de riesgo epidemiológico de transmisión de enfermedades [5]. Al finalizar, se determina en palabras de los autores, que es posible identificar patrones dentro del análisis de los resultados obtenidos en los clústeres, por medio de gráficos de comportamiento de las variables con respecto a las demás, donde tras analizarlas, se toman decisiones basadas en los resultados para crear programas de prevención y promoción [5].

4.2 Marco Conceptual

Para comprender la importancia de la extracción de información de los datos generados por las diferentes acciones que realizan los seres humanos es necesario comprender los conceptos claves:

4.2.1 Log

Un log es un registro de actividad de un sistema, que generalmente se guarda en un fichero de texto, al que se le van añadiendo líneas a medida que se realizan acciones sobre el sistema. Se utiliza en muchos casos distintos, para guardar información sobre la actividad de sistemas variados. A través de los log se puede encontrar información para detectar posibles problemas en caso de que no funcione algún sistema como debiera o se haya producido una incidencia de seguridad [6].

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Para el presente trabajo se utilizará un log de acceso, que al ser analizado da como resultado información acerca del tráfico del sitio.

4.2.2 Log de Servidores

Consiste en uno o más ficheros de texto automáticamente creados y administrados por un servidor, en donde se almacena toda la actividad que se hace sobre éste. Cada servidor, dependiendo de su implementación y/o configuración, podrá o no crear determinados Logs.

4.2.3 Log de Acceso

Es un tipo de log de servidor. En un servidor web, un log de accesos es un archivo o grupo de archivos que contiene una lista de cada archivo que fue accedido en el servidor.

En general, se almacena la dirección IP, la fecha y hora de acceso, el archivo accedido, el navegador o agente usado, el sistema operativo y versión, la URL de referencia, bytes servidos, etc. Cabe resaltar que existen múltiples formas en que se puede almacenar esta información, dependiendo del software utilizado para correr el servidor. Entre estas formas, la más popular es el formato estándar recomendado por la W3C.

El log de accesos puede servir, entre otras cosas, para crear las estadísticas de acceso a un sitio web, así como para determinar cuándo un servidor web se sobrecarga.

También existen implementaciones que separan en diferentes archivos el log de accesos, por ejemplo, un archivo exclusivo para los accesos que dieron error (el log de errores).

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

4.2.4 Dirección IP

Las direcciones IP (IP es un acrónimo para Internet Protocol) son un número único e irrepetible con el cual se identifica una computadora conectada a una red que corre el protocolo IP.

Una dirección IP (o simplemente *IP* como a veces se les refiere) es un conjunto de cuatro números del 0 al 255 separados por puntos. En realidad, una dirección IP es una forma más sencilla de comprender números muy grandes [6].

4.2.5 Analog

Es un analizador de los ficheros log generados por los servidores web. Genera automáticamente una página web XHTML con el resultado del análisis, que puede configurarse con todo detalle.

La documentación está en inglés, pero el programa crea informes en más de 30 idiomas, incluido el castellano. Además, no necesita instalación: hay que descomprimir el archivo zip en una carpeta y ejecutar el programa `Analog.exe`.

4.2.6 Weka

(WaikatoEnvironmentforKnowledgeAnalysis) - Entorno para Análisis del Conocimiento de la Universidad de Waikato) es un conocido software para aprendizaje automático y minería de datos escrito en Java y desarrollado en la Universidad de Waikato. WEKA es un software libre distribuido bajo licencia GNU-GPL [7].

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Ventajas de Weka:

- Está disponible libremente bajo la licencia pública general de GNU.
- Es muy portable porque está completamente implementado en Java y puede correr en casi cualquier plataforma.
- Contiene una extensa colección de técnicas para pre-procesamiento de datos y modelado.
- Es fácil de utilizar por un principiante gracias a su interfaz gráfica de usuario [8].

Weka soporta varias tareas estándar de minería de datos, especialmente, pre-procesamiento de datos, clustering, clasificación, regresión, visualización, y selección. Todas las técnicas de Weka se fundamentan en la asunción de que los datos están disponibles en un fichero plano (flat file) o una relación, en la que cada registro de datos está descrito por un número fijo de atributos (normalmente numéricos o nominales, aunque también se soportan otros tipos). Weka también proporciona acceso a bases de datos vía SQL gracias a la conexión JDBC (Java DatabaseConnectivity) y puede procesar el resultado devuelto por una consulta hecha a la base de datos. No puede realizar minería de datos multi-relacional, pero existen aplicaciones que pueden convertir una colección de tablas relacionadas de una base de datos en una única tabla que ya puede ser procesada con Weka [9].

4.2.7 Webmining

El Webmining es una metodología de recuperación de la información que usa herramientas de la minería de datos para extraer información tanto del contenido de las páginas, de su estructura de relaciones (enlaces) y de los registros de navegación de los usuarios [10].

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Webmining es la aplicación de técnicas de minería de datos para descubrir los patrones de la Web. De acuerdo a los objetivos de análisis, la minería web se puede dividir en tres tipos diferentes, que son la minería de uso Web, Web Content minería y estructura de minería de la Web [11].

4.2.8 Minería de Datos (Data Mining)

La minería de datos (es la etapa de análisis de "KnowledgeDiscovery in Databases" o KDD), es un campo de las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos. Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos. El objetivo general del proceso de minería de datos consiste en extraer información de un conjunto de datos y transformarla en una estructura comprensible para su uso posterior.

Además de la etapa de análisis en bruto, que involucra aspectos de bases de datos y gestión de datos, procesamiento de datos, el modelo y las consideraciones de inferencia, métricas de Intereses, consideraciones de la Teoría de la complejidad computacional, post-procesamiento de las estructuras descubiertas, la visualización y actualización en línea [12].

4.3 Estado del Arte

4.3.1 A Nivel Internacional

“Aplicación de minería de datos para determinar patrones de consumo futuro en clientes de una distribuidora de suplementos nutricionales” es el proyecto llevado a cabo por Miguel Ángel Grández Márquez para optar el título de Ingeniero Informático de Sistemas de la Universidad San

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Ignacio de Loyola de Lima, Perú en 2017, su tesis está dirigida a la tienda Pardo Miraflores, se plantea como principal objetivo la identificación de los compradores potenciales y caracterización de los mismo de acuerdo a sus características demográficas, historial de transacciones y modelo de navegación. Una vez realizado el estudio, Grández concluye que el producto más demandado corresponde a la Proteína Néctar Isolate y además pudo segmentar los clientes de la tienda Pardo Miraflores, de acuerdo a características tales como el género, la edad y el deporte realizado por cada uno de los usuarios, con el fin de conocer el perfil de los compradores de cada uno de sus productos y generar estrategias para el incremento de los ingresos del comercio [13].

Un año más tarde, en el 2018, es publicado un informe por Esther Ruiz Lobaina y Pedro Romero Suárez, el cual tuvo como objetivo “mejorar la calidad de la información que pertenece a la base de datos CubaCiencia, del Instituto de Información Científico y Tecnológico” [14], este proyecto se llevó a cabo a través de la aplicación de una herramienta de la minería de datos, denominada Rapid Miner v.5.2. A su vez, se usó la técnica de árbol de decisión, debido a que se quería conocer las mayores búsquedas o temáticas más consultadas en cada uno de los cuatro segmentos de las ciencias que se pueden identificar en la base de datos, los cuales son: las Ciencias Agropecuarias, las Ciencias Biomédicas, Técnicas y Ciencias Sociales. Asimismo, se empleó la la Matriz de Correlación, el Scatter Plot 3D, Desviación Standard y Surface 3D para analizar los datos y poder visualizarlos a través de gráficas, todo esto, con la intención de mejorar el servicio y la información por parte de la base de datos, para que todo usuario pueda acceder y encontrar fácilmente su motivo de consulta. Finalmente, se logró estandarizar la información, mejorar el sistema de gestión bibliotecario y se creó un repositorio para realizar vigilancia y prospección, debido a los patrones obtenidos al finalizar el estudio de minería de datos.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

En lo referente a México, Álvaro Rodríguez, Director de operaciones, a través de una página de consultoría, plantea la importancia de la minería de datos y la aplicación de sus tecnologías para la toma de decisiones de cualquier organización, de igual manera, expone, por medio de su experiencia, la utilidad que le ha dado a un modelo informático desarrollado en su empresa, puesto que “brinda una integración y análisis de información en tiempo real de toda la información que integra el expediente de un trabajador” [15]. Una vez aplicado este modelo, será más fácil para el informante y consultante mantener una mejor comunicación, ya que los datos se actualizan constantemente, lo que se traduce en un beneficio mutuo debido a la aplicación de dichas tecnologías.

4.3.2 A Nivel Nacional

En el 2015, de acuerdo al repositorio de la Universidad de Envigado, Felipe Cálad Noreña, Ingeniero de Sistemas y Computación de la Escuela de Ingeniería de Antioquia, tuvo como idea realizar la segmentación automatizada de clientes a partir de técnicas de minería de datos, la aplicación del estudio se realiza a la empresa Tiendacol S.A., una empresa perteneciente al sector de la moda de la ciudad de Medellín cuya actividad es el comercio de prendas de vestir y accesorios. En su trabajo se propone como objetivo principal “Diseñar una segmentación de clientes automatizada en la empresa Tiendacol S.A. mediante minería de datos y la técnica k-means clustering, que permita clasificar sus clientes basado en las características demográficas y los patrones de compra de los mismos” [16]. Luego de realizar los estudios y análisis correspondientes, Felipe concluye que fue posible clasificar los clientes de la empresa Tiendacol S.A. en tres grandes grupos, de acuerdo con la frecuencia y cantidad de dinero que gastan en las compras dichos clientes. Es así, como se identifica la necesidad de generar estrategias tanto de

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

fidelización para aquellos clientes que se encuentran en el grupo de los mejores clientes, como de captación para quienes han realizado compras ocasionales o aún no conocen la tienda. Es importante resaltar que el proyecto se basa en la aplicación de diversas técnicas de minería de datos. En palabras del autor “mediante la integración de minería de datos con sistemas de inteligencia de negocios se pueden construir herramientas muy poderosas para el apoyo de la toma de decisiones de toda empresa” [16].

Por otra parte, en la capital del país, Laura Caro, a través de su informe “Identificación de las ventajas de aplicar minería de datos en el negocio automotriz” publicado en 2016 plantea el motivo de su proyecto, el cual se basa en conocer la viabilidad de la información en los datos de un concesionario y la información oculta a través de la aplicación de herramientas de minería de datos. Para llevar a cabo el estudio, se plantearon las variables relevantes y su caracterización, las variables a analizar fuer el mes, año de venta, concesionario, tipo de compra, precios de venta, vendedor, cliente, uso, grupo, tipo, marca, gama, referencia, color y modelo. Posteriormente, a través del software SPSS se continuó con el estudio y análisis de las variables, obteniendo datos y gráficos estadísticos que finalmente llevaron a concluir que las mayores ventas realizadas corresponden a las de 4 referencias (HFC1063K, HFC1035K, HFC1035KD y HFC1061L), lo que indica que “la línea de vehículos con una mayor proporción en ventas es la HFC, al igual que la aceptación de los vehículos 1035K y su variante el 1935KD” [17], esta conclusión permite a su vez, que la empresa analice las características de la referencia más vendida para poder incrementar las ventas de aquellas que tienen menos demanda, ajustándose a las necesidades y preferencias del cliente.

A su vez, en el 2019, se realiza una publicación en la Revista Politécnica, la cual “es una publicación académica semestral de ciencia, tecnología e innovación del Politécnico Colombiano

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Jaime Isaza Cadavid, Medellín, Colombia” [18]. Cabe resaltar que está dirigida a la comunidad académica y científica nacional e internacional. Por ende, a través de la publicación “Minería de datos educativos: análisis del desempeño de estudiantes de ingeniería en las pruebas saber-pro” realizada por Ana Isabel Oviedo Carrascal y Jovanny Jiménez Giraldo, Doctora en Ingeniería Electrónica e Ingeniero Electrónico, ambos docentes de universidades (Universidad Pontificia Bolivariana, Medellín y Universidad Católica de Oriente, Rionegro), se realiza un estudio de las Pruebas Saber Pro, analizando diferentes variables, algunas de ellas académicas y económicas, así como los resultados que obtuvieron los estudiantes de Ingeniería de Antioquia en estas pruebas, presentadas en el año 2016 por los mismos. A lo largo del documento se hace énfasis en las tres técnicas de minería de datos aplicadas para el estudio, las cuales fueron, en primer lugar, el análisis de clúster, el sistema de votación de varias técnicas y la predicción del desempeño en la prueba, acorde a las variables seleccionadas anteriormente. Cabe resaltar, que, durante el desarrollo de su estudio, aplicaron la metodología CRISP-DM a través de sus seis fases, las cuales se desglosan en el entendimiento del negocio, de los datos, preparación de los datos, modelado, evaluación y, por último, la implementación. De un total de 49.021 registros, se obtuvo 108 variables con información de diversos ámbitos de los estudiantes de Ingeniería, así como los resultados individuales de los módulos de dicha prueba. En cuanto al tratamiento de los datos se llevó a cabo a través de la herramienta Weka 3.8, de esta manera se generaron 6 clústeres, donde se clasificó a los estudiantes de acuerdo al desempeño obtenido, reconociendo su género, estrato, personas a cargo, entre otros factores claves para poder caracterizar a los estudiantes. Tras utilizar K-means en el análisis de clúster, fue posible detectar que, en su mayoría, los estudiantes que obtienen destacables resultados en inglés, poseen mejores condiciones económicas que el resto, por su parte, aquellos que tienen buen desempeño en lectura crítica son personas con menos recursos y, por

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

ende, el pago de su matrícula es menor. Finalmente, a partir de 26 variables seleccionadas en el experimento, entre las cuales se destacaron algunas como el número de personas a cargo, método de enseñanza, el carácter académico de la institución y la economía en su hogar, se logró predecir el desempeño en las pruebas con una exactitud del 81% [18].

4.3.3 A Nivel Regional

De acuerdo con Carlos Alberto Ortiz Cárdenas, Ingeniero Industrial de la Universidad Santo Tomás, quien realizó el estudio en el año 2015 denominado “Caracterización de los municipios del departamento de Santander usando indicadores de gestión municipal mediante técnicas de minería de datos”, el uso que le da a la minería de datos en su proyecto es para cumplir con el objetivo de caracterizar los diferentes municipios de Santander, compararlos entre sí y establecer cuáles presentaban una mayor eficiencia en el manejo de sus recursos financieros, esto, a través de la aplicación del análisis de clúster y la metodología KDD. Finalmente, Carlos pudo establecer diversas conclusiones, algunas de ellas, relacionadas con aquellos departamentos que deben ser tomados como ejemplo por sobresalir en cuanto a desarrollo y gestión administrativa pública (Barrancabermeja, Piedecuesta, San Gil, Bucaramanga y Floridablanca), y otras con relación a la verificación del personal, reducción de costos, realizar mayores inversiones y mayores entes de control, representando de esta manera, una ventaja para cada uno de los municipios al conocer los datos que arrojó el estudio y poder mejorar considerablemente frente a su situación actual por el bien de la comunidad [19].

Por otra parte, en la Revista UIS de Ingenierías, se publicó en 2016 “Perfil de la investigación sobre inteligencia de negocios en América Latina” por A. Bustamante, E. Galvis y L.C. Gómez, tres jóvenes pertenecientes al Grupo STI, en su estudio se presenta un análisis en el

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

periodo entre el 2000 y 2012, donde se obtuvo la información de la base de datos SCOPUS, “adicionalmente, se evidencia que los temas más investigados están relacionados con la minería de datos y que las instituciones de Brasil producen la mayor cantidad de la literatura en la región, pues el 56% de las publicaciones provienen de ese país” [20]. Para el desarrollo de su estudio se implementó una metodología compuesta de 9 pasos, los cuales, a su vez, están agrupados en tres fases: inteligencia, análisis y diseño, y selección. Lo primero que se pudo identificar mediante la investigación fue el crecimiento de las publicaciones realizadas en América Latina, donde en el año 2000 se tenían tan solo 20 publicaciones, mientras que para el año 2012 resultaron 261, lo que equivale a un crecimiento del 1086% aproximadamente. Sin embargo, cabe resaltar que los países que incrementaron sus publicaciones considerablemente fueron Brasil, México y Chile, mientras que, por ejemplo, en Colombia, no hubo gran diferencia durante este periodo analizado [20].

De acuerdo a la Revista Colombiana de Computación, en el año 2019 es publicado el artículo “Aplicación para la gestión y el análisis de información relacionada con la deserción estudiantil universitaria”, en el cual los autores, Luz Yamile Caicedo Chacón, Cristian Noé Cárdenas Parra, Juan Sebastián Müller Rueda y Jeniffer Tatiana Ortiz Bernal, ingenieros de sistemas, investigadores de tiempo completo y desarrolladores de software de la Fundación Universitaria San Gil, proponen la implementación y el desarrollo de un software bajo la metodología SCRUM, en busca de tener un registro y control de los diferentes datos personales, ya sean socioeconómicos, psicosociales o aquellos generados a lo largo de la estadía de los estudiantes en las Instituciones de Educación Superior, todo esto a través de técnicas de minería de datos, para determinar patrones que permitan identificar la existencia del riesgo o posible deserción a través de alarmas. Para el desarrollo de su propuesta se aplicó en primer lugar una encuesta de deserción a todos los estudiantes de las tres sedes de la universidad, las cuales fueron

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

posteriormente analizadas a través de SPSS y la herramienta Weka, en la que se aplicó específicamente dos técnicas, K-means y J48. Finalmente, diseñada y construída la herramienta software que permite recolectar información personal, académica, familiar y financiera de los estudiantes; se planteó que los datos obtenidos son estudiados mediante técnicas de minería de datos con el fin de generar reportes estadísticos y nuevas estrategias o programas de acompañamiento que apoyen los procesos de permanencia y graduación de los estudiantes de UNISANGIL [21].

5. Metodología

La metodología implementada en el desarrollo del presente proyecto está directamente asociada a la minería de datos, la cual hace referencia a un conjunto de algoritmos que permiten la identificación de patrones útiles y novedosos que se encuentran “ocultos” en grandes bases de datos. Por tal razón es importante aclarar los 4 pasos contemplados por el KDD que serán desarrollados y expuestos posteriormente:

1. Entendimiento del problema: Se define la pregunta que se quiere resolver y se establecen los objetivos específicos a trabajar.
2. Selección de datos: Se determina qué bases de datos pueden ser utilizadas para analizar el problema y se consolidan los conjuntos de información. Específicamente se trata del log del servidor del 06 al 27 de febrero del 2020, información analizada a través del software Weka y Analog para la caracterización de los usuarios de la empresa Adonai Employment.
3. Pre-procesamiento y limpieza: Se busca manejar los datos faltantes junto con los valores atípicos que se identifiquen en las bases de datos.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

4. Transformación de datos: Consiste específicamente en la búsqueda y hallazgo de patrones de los usuarios de la empresa Adonai Employment más representativos, tomando la información acumulada hasta este momento [22].

5.1 Tipo de Investigación

Tras haber analizado diferentes criterios que permiten realizar la clasificación de la presente investigación, se determinó el objetivo general como el criterio a utilizar, puesto que, en la investigación holística, se define el tipo de investigación de acuerdo al verbo de dicho objetivo general del proyecto, asimismo, el tipo de investigación está directamente relacionado con el logro de conocimiento que se quiere alcanzar.

Por consiguiente, de acuerdo al nivel de profundidad expresado en el objetivo general, se determina que la investigación es descriptiva, debido a que está dirigida a detectar la tendencia y comportamiento de los usuarios de la empresa Adonai Employment. En palabras de Sampieri, “con los estudios descriptivos se busca especificar las propiedades, las características y los perfiles de personas, grupos, comunidades, procesos, objetos o cualquier otro fenómeno que se someta a un análisis” [23]. Es decir, se describen tendencias de un grupo o población, el cual, en este caso específico corresponde a todas las personas que ingresan al sitio web de Adonai Employment.

Del mismo modo, Jacqueline Hurtado, en su libro Metodología de la Investigación Holística, plantea que “la investigación descriptiva consiste en identificar las características del evento en estudio” [24].

Tras complementar los dos puntos de vista expuestos anteriormente, los cuales están totalmente relacionados y expresan claramente las características del tipo de investigación, fue

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

posible determinar que se trata de una investigación descriptiva, dado que a lo largo del documento se ha hecho énfasis en que la finalidad es conocer los patrones de comportamientos de los usuarios de en cuanto al acceso a la web, de manera que se determinen las tendencias y características de los clientes junto con los servicios más solicitados, para finalmente, establecer estrategias que permitan crecer como empresa sin ánimo de lucro, abarcando mayores usuarios y territorios en busca de una mano ayuda.

5.2 Diseño de la Investigación

Los datos del proyecto de investigación se recolectarán a través de un diseño documental, este diseño documental se trata de una investigación secundaria, puesto que la fuente es específicamente la base de datos de la empresa Adonai Employment. Para realizar el análisis de los datos generados por el acceso de los usuarios a la página www.adonai-employment.org se procede a escoger el log de acceso que se va a utilizar durante el análisis, en este caso se seleccionó el log de los días 06 al 27 de febrero del año en curso.

En primera instancia, se obtendrá el log denominado “log sucio” y este será el utilizado para generar los reportes e informes con el software libre Analog 6.0. Estos datos son recolectados al efectuar el desarrollo del proyecto, debido a que se necesita esta información para poder cumplir con cada uno de los objetivos propuestos. La información que se obtiene constituye a información de períodos pasados, específicamente a las tres últimas semanas del mes de febrero, en conclusión, se trata de una investigación con diseño retrospectivo evolutivo, ya que el análisis se desarrolla a lo largo del tiempo.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Para generar los informes con Analog es necesario crear un back up del log sucio modificándole su extensión original (.txt) por “.log”, esta extensión es necesaria para el análisis del log. Posteriormente, se procede a copiar el log con la extensión .log a la carpeta que contiene el Analog (estos es un requisito para el correcto funcionamiento del programa).

A su vez, se abre el archivo Analog.cfg para darle los parámetros necesarios para el análisis del log, como el idioma la ruta del log y el formato que este tiene.

Algunas de las variables que serán tratadas son el número de peticiones por mes que tuvo el servidor, identificando el o los meses donde hubo mayor tráfico en la página, además, también se analizará la actividad diaria (de domingo a sábado) de manera que se evidencie el día en el que la página web recibe más visitas en comparación con los demás. A su vez, la actividad total por cada hora de la semana (de 0 a 23), de manera que se pueda analizar la hora con mayor ingreso de usuarios. Por otra parte, se tendrá información acerca de las organizaciones o personas que hicieron mayores consultas en el sitio, teniendo como resultado los IP.

Con respecto a la cantidad de datos a recoger, debe aclararse que toda la información se obtiene de la base de datos de Adonai Employment, la cual va a tener almacenados todos los datos del sistema de la empresa. No obstante, la información o datos a utilizar son los generados a lo largo del 06 al 27 de febrero del 2020, como se mencionaba anteriormente.

Tras haber ingresado al log del servidor, lo primero que se realiza es el estudio del log sucio del servidor encontrado, el cual es analizado posteriormente con ayuda de la herramienta Analog para obtener la información del uso de los servidores de la página a través de un análisis detallado de los registros. Una vez analizado, se procede a limpiar el log y se preparan los datos o información extraída.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Seguido a la limpieza, se ubica el proceso de parametrización del log, donde se realizan los procedimientos necesarios por medio de la herramienta Excel hasta llegar a parametrizar todos los atributos del log.

Después de parametrizar el log por medio de la herramienta Excel, se toman estos datos y se realiza un archivo con formato Arff, el cual es interpretado por el software libre Weka para aplicar el algoritmo A priori, lo que genera las reglas de asociación para predecir el comportamiento de los usuarios. De esta manera se obtienen gráficos estadísticos que muestran sus diferentes atributos, como, por ejemplo, hora y fecha de entrada, pertinentes para crear patrones de comportamiento acerca de las tendencias de la actividad de los usuarios en la página web de Adonai Employment. Por último, se evalúa si el modelo responde a los objetivos planteados inicialmente en el proyecto y se sacan conclusiones de acuerdo a los informes arrojados por el software libre Analog, así como los datos estadísticos ofrecidos en Weka.

5.3 Método

Tabla 1. *Método de Investigación*

Objetivo Específico	Método	Actividades
Describir los patrones de comportamiento de navegación de los usuarios de la empresa Adonai Employment utilizando el software Analog.	Obtención y generación de reporte a través de gráficos estadísticos que muestran los diferentes atributos del log utilizando el software Analog.	<ul style="list-style-type: none"> - Crear un back up del log sucio modificándole su extensión original (.txt) por “.log”. -Generar lo estadísticos de navegación de los usuarios por la web de la empresa a través de la aplicación del software Analog. -Analizar e interpretar los datos arrojados al utilizar el log del servidor a través del Analog

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Implementar una sábana de datos con el log del servidor de la empresa Adonai Employment por medio de un data warehouse para conocer las tendencias de navegación y características de los usuarios utilizando reglas de asociación.	Aplicación del algoritmo A priori para generar las reglas de asociación a través del software Weka.	<ul style="list-style-type: none"> -Descargar el log de la página web de la empresa Adonai Employment. -Limpiar y parametrizar el log. -Realizar e interpretar el archivo con formato Arff para predecir el comportamiento de los usuarios.
Validar los resultados de navegación del usuario generados por el software Analog con los resultados del software Weka.	Revisión de la precisión de los datos arrojados tanto en los algoritmos del software Weka como en los informes del software Analog.	<ul style="list-style-type: none"> - Interpretar los datos y modelado predictivo a través del software Weka. -Detallar la información contenida en los informes desarrollados a través del Analog. -Analizar el manejo de la información proporcionada por la empresa y corregir los errores que puedan afectar el buen desempeño de la misma.

6. Aplicación de las Herramientas de Web Mining para la Interpretación de la Información de Log de Acceso

Para realizar el análisis de los datos generados por el acceso de los usuarios a la página www.adonai-employment.org se escogió el log de acceso de los días 7 a 27 de febrero del año en curso.

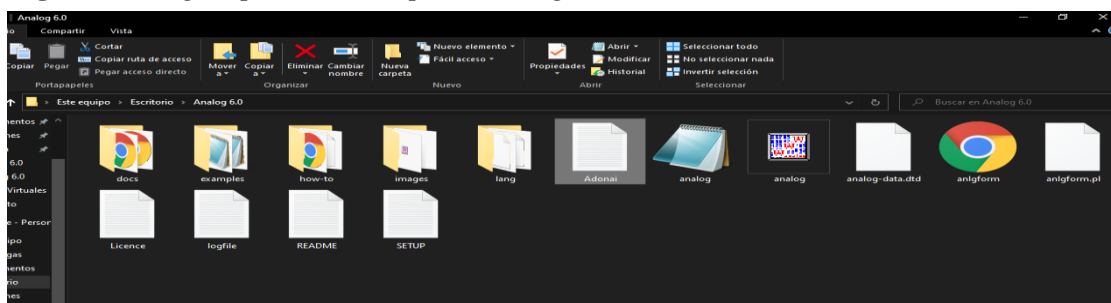
Estos datos pertenecen al log de un servidor, cuya información es confidencial y al ser tratada correctamente fue posible extraer patrones o reglas de comportamientos de usuarios de Adonai Employment.

7. Resultados

7.1 Reportes e Informes con Analog

Para la generación de reportes con el software Analog se creó una copia de seguridad del bloc de notas del log sucio y a su vez, se modificó su extensión, por lo cual pasó de ser .txt a .log. Seguidamente el log con extensión .log se copió en la carpeta denominada “Analog” como se muestra en la siguiente imagen. Este fue uno de los requisitos para conseguir el correcto funcionamiento del programa.

Figura 2. Log Copiado en Carpeta Analog 6.0.



Luego se abrió el archivo Analog.cfg al que se le dieron los parámetros de configuración necesarios para el análisis del log como el idioma, la ruta del log y el formato que este tiene.

Una vez abierto el Analog.cfg, se editaron los siguientes campos: %s (numerical IP address of client), %d (day), %M (month in digits), %Y (year, last two digits), %h (hour of the day), %n (minute of the hour), %j (seconds).

Para el presente caso el LOGFORMAT fue T (%s - [%d/%M/%Y:%h:%n:%j] "%r /%q %j" %c %b %B), donde adicionalmente se encuentran los campos correspondientes a archivos requeridos, extensiones de los mismos, el código de estado (http), número de bytes transferidos y finalmente el navegador a través de cuál se exploró.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

La figura 4 contiene los reportes globales como son las peticiones exitosas que se realizaron con un valor de 732 durante las 3 semanas analizadas; por otro lado, las peticiones no logradas que corresponden a 83; donde se puede inferir que aproximadamente el 10% de las peticiones realizadas por los usuarios fueron fallidas.

A continuación, se muestra el resumen de las peticiones realizadas cada día de la semana, a través de esta tabla se puede analizar que el día con mayor flujo o movimiento fue el martes de acuerdo a las 139 peticiones que se recibieron, seguidas por el día miércoles con 131 peticiones. En contraste, el domingo fue el día que menos tránsito tuvo la página de Adonai Employment con tan solo 75 peticiones.

Figura 5. Resumen Diario Adonai Employment.

Resumen diario		
(Ir a: Armba Resumen general Informe mensual Resumen diario Resumen horario Informe de dominios Informe de organización Info		
<i>Este reporte enlista la actividad total por cada día de la semana, sumados por todas las semanas en el reporte.</i>		
Cada unidad (■) representa 1 petición de una página.		
día	No. pet.	Págs.
Dom	75	0
Lun	93	0
Mar	139	0
Mié	131	0
Jue	113	0
Vie	104	0
Sáb	77	0

Por su parte, el resumen de horario se especifica a continuación en la figura 6, en la que se puede observar que la hora con mayor tráfico fue las 10 am con 76 peticiones, seguida de las 3 am con 57 de las 732 recibidas durante el periodo del 6 al 27 de febrero del 2020. Asimismo, la hora con menor flujo fue la 1 am (tan solo 11 peticiones).

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Figura 6. Resumen Horario Adonai Employment.

Nota: Adaptado de Log Adonai Employment (2020)

7.1.1 Informe de Organización

En este informe se detallan las organizaciones de los computadores que pidieron determinado archivo. Además, en la gráfica se observan las diferentes organizaciones que ingresaron a la página, asignándole un color a cada una de ellas.

En la tabla de la derecha se observa de manera más detallada la lista de organizaciones que ingresaron, junto con el número de peticiones realizadas y la cantidad de bytes transferidos por cada una de ellas.

Figura 7. Informe de la Organización Adonai Employment.

Nota: Adaptado de Log Adonai Employment (2020)

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Tras verificar la información aportada por el Software Analog, se pudo deducir que la subred o familias de IP que realizaron mayor acceso al servidor de la empresa fue la 207.46.

7.1.2 Informe de Sistemas Operativos

En este informe se detallan los diferentes sistemas operativos junto con el número de peticiones realizadas. De esta manera, se evidencia que el sistema operativo más usado es Windows, descartando su versión. Pese a esto, Windows 10 fue el sistema operativo más utilizado por los visitantes de la página de Adonai Employment.

Tabla 2. Listado de los sistemas operados ordenados por número de peticiones por páginas.

NO.	NO. PET.	PÁGS.	SO
1	5	0	Unix
	5	0	Linux
2	43	0	Robots
3	54	0	Macintosh
4	319	0	Sistema Operativo desconocido
5	311	0	Windows
	156	0	Windows 10
	134	0	Windows desconocido
	1	0	Windows 8
	2	0	Windows 8.1
	2	0	Windows 7
	3	0	Windows XP
	13	0	Windows ME

Nota: Adaptado del Log Adonai Employment (2020).

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

7.1.3 Informe de Códigos de Estado

En este informe se obtienen los códigos de estado de HTTP de las peticiones realizadas. A continuación, se detallan los códigos de estado que aparecen en el log a través de la gráfica:

Figura 8. Informe de Códigos de Estado.



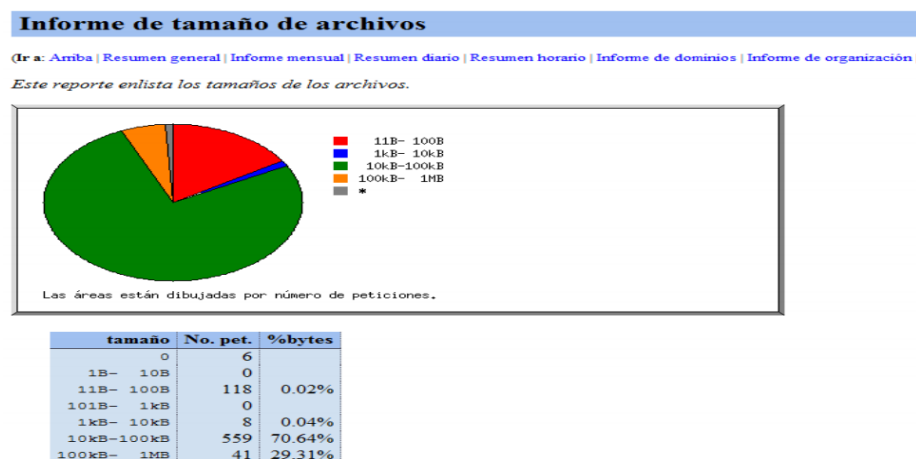
Nota: Adaptado de Log Adonai Employment (2020)

En la gráfica y su respectiva tabla se visualizan los códigos de estado del log, entre los que predomina el código 202, es decir el de acceso correcto (OK) con un total de 732 peticiones. Lo que quiere decir que, la gran mayoría de las peticiones no presentó ningún tipo de error y, por el contrario, obtuvo una respuesta exitosa.

Por otro lado, el código de estado HTTP 404 indica que la petición a la URL solicitada por el usuario no ha sido modificada desde la última vez que fue requerida, es decir, que el documento no fue encontrado con exactamente 83 peticiones.

7.1.4 Informe de Tamaño de Archivos

A continuación, se detalla el tamaño de los archivos requeridos por los usuarios que ingresaron a esta página, lo que permitió identificar que el tamaño más requerido osciló entre los 10kB y los 100kB con 559 peticiones y un porcentaje de bytes de 70,64%.

Figura 9. Informe de Tamaños de Archivos.

Nota: Adaptado de Log Adonai Employment (2020)

7.1.5 Informe de Peticiones

En la gráfica se observan los archivos que se encuentran en el sitio, donde gran parte de los archivos pedidos fueron de tipo get.

Figura 10. Informe de Peticiones.

Nota: Adaptado de Log Adonai Employment (2020)

Por otra parte, en la tabla 3 es posible observar de manera más explícita los archivos más solicitados, estos son aquellos que representan por lo menos 20 peticiones y se ordenaron de

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

manera descendente. Además, se incluyó el % de bytes transferidos, la última fecha en la que se hizo la petición, y el tipo de archivo más solicitado.

Tabla 3. Listado de archivos que representan por lo menos 20 peticiones

Pet.	%bytes	Última hora	Archivo
724	100%	27/Feb/2020 12:59 pm	Get
115	0.02%	27/Feb/2020 12:59 pm	get?robots.txt
45	4.34%	27/Feb/2020 2:03 am	get?meet_our_team
34	5.20%	27/Feb/2020 2:02 am	get?contact_uslocation
27	1.76%	27/Feb/2020 2:02 am	get?about_us
26	1.69%	27/Feb/2020 2:03 am	get?home
26	1.64%	27/Feb/2020 2:03 am	get?employer_services
25	1.44%	27/Feb/2020 2:03 am	get?job_seeker_services
24	1.41%	27/Feb/2020 2:03 am	get?resourceslink
22	1.20%	27/Feb/2020 2:03 am	get?success_stories
21	7.24%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/Charles_Wickens_Picture1.11784601_std.jpg
21	14.00%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/Lorie.17275508_std.jpg
20	5.82%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/Jose39_1.16792659_std.jpg
20	7.09%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/DUC_0012_copy_copy_158144156.16790052_std.jpg
19	14.41%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/Kyle.17192129_std.jpg
19	6.30%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/Debbie101_1.91142440_std.jpg
17	4.82%	26/Feb/2020 1:35 pm	get?yahoo_site_admin/assets/images/TOM53_1.91135850_std.jpg
8		25/Feb/2020 9:46 pm	[no listados: 2 archivos]

8. Limpieza del Log

Para llevar a cabo el análisis de los datos a través de la herramienta Weka se hizo necesario realizar una “limpieza” o filtro de la información suministrada en log. Una gran ayuda para realizar este proceso fue el reporte generado por el Analog referente a los tipos de archivos solicitados, ya que las extensiones que presentaron menos solicitudes no deben ser consideradas como relevantes dentro del análisis del log.

Para realizar la limpieza, se introdujeron los datos en la herramienta ofimática Excel, la cual brindó facilidades al momento de realizar esta labor, a través de un filtro de información.

En primer lugar, se eliminaron las líneas cuyas extensiones eran .gif, puesto que esta extensión hace referencia a archivos imágenes, las cuales no representan una fuente de información. A su vez, fueron eliminadas las líneas con extensiones .jpg y .css. De esta manera se obtuvo como resultado el log limpio que se muestra a continuación.

Asimismo, es válido aclarar que no bastó con la limpieza de los campos innecesarios, sino que también se removieron los registros del error producidos por diferentes causas como el conocido 404 NOT FOUND o mensaje de página no encontrada.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Figura 11. Log Limpio Adonai Employment.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	72.131.38.53	7-feb-20	13	40	24	800	200	13465	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
2	72.131.38.53	7-feb-20	13	42	15	800	200	25678	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
3	76.199.174.27	7-feb-20	14	14	37	800	200	15158	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	www	adonai-employment.org	
4	76.199.174.27	7-feb-20	14	14	39	800	200	22467	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	www	adonai-employment.org	
5	76.199.174.27	7-feb-20	14	14	40	800	200	80325	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
6	76.199.174.27	7-feb-20	14	14	40	800	200	77345	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
7	76.199.174.27	7-feb-20	14	14	40	800	200	66083	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
8	76.199.174.27	7-feb-20	14	14	40	800	200	155334	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
9	76.199.174.27	7-feb-20	14	14	40	800	200	67880	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
10	76.199.174.27	7-feb-20	14	14	40	800	200	166532	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
11	76.199.174.27	7-feb-20	14	14	40	800	200	82569	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	adonai-employment.org		
12	76.199.174.27	7-feb-20	14	16	37	800	200	35678	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	www	adonai-employment.org	
13	76.199.174.27	7-feb-20	14	16	53	800	200	22467	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	www	adonai-employment.org	
14	76.199.174.27	7-feb-20	14	30	7	800	200	35678	http://www.adonai-employment.org/	Mozilla/compatible;_MSIE_	www	adonai-employment.org	
15	75.17.150.81	7-feb-20	18	4	20	800	200	22467	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
16	75.17.150.81	7-feb-20	18	4	21	800	200	103334	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
17	75.17.150.81	7-feb-20	18	4	21	800	200	77345	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
18	75.17.150.81	7-feb-20	18	4	21	800	200	80325	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
19	75.17.150.81	7-feb-20	18	4	24	800	200	82569	http://adonai-employment.org/	Mozilla/Windows;_U_	Win	adonai-employment.org	
20	193.47.80.146	8-feb-20	2	28	35	800	200	48	http://www.exalead.com/	Mozilla/compatible;_Exabi_	adonai-employment.org		
21	193.47.80.146	8-feb-20	2	28	36	800	200	77345	http://www.exalead.com/	Mozilla/compatible;_Exabi_	adonai-employment.org		
22	67.51.233.200	8-feb-20	17	54	33	800	200	22467	http://www.google.com/url?sa=t	Mozilla/Macintosh;_U_	Int	adonai-employment.org	
23	67.51.233.200	8-feb-20	17	54	42	800	200	66083	http://adonai-employment.org/	Mozilla/Macintosh;_U_	Int	adonai-employment.org	

9. Parametrización del Log

Tabla 4. IP Usuarios Log

IP	
Clase A	0-127
Clase B	128-191
Clase C	192-255

Tabla 5. Fecha de Acceso Log

FECHA	
07-feb-20	0
08-feb-20	1
09-feb-20	2
10-feb-20	3
11-feb-20	4
12-feb-20	5
13-feb-20	6
14-feb-20	7
15-feb-20	8
16-feb-20	9
17-feb-20	10
18-feb-20	11
19-feb-20	12
20-feb-20	13

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

21-feb-20	14
22-feb-20	15
23-feb-20	16
24-feb-20	17
25-feb-20	18
26-feb-20	19

La hora, los minutos y los segundos no fueron parametrizados, ya que así se genera una mejor precisión al momento de realizar algún requerimiento.

El código de estado del log siempre era 200, por ello se deja de la misma manera.

Tabla 6. *Bytes Transferidos*

BYTES	
0-50.000	0
50.001-100.000	1
100.001-150.000	2
150.001-200.000	3

Tabla 7. *Navegador Utilizado Log*

NAVEGADOR	
Mozilla Firefox	0
Chrome 86	1
Chrome 78	2
Opera 15	3

Tabla 8. *URL de Búsqueda Log*

URL	
adonai-employment.org/	0
http://adonai-employment.org	1
http://adonai-employment.org/	2
http://adonai-employment.org/about_us	3
http://adonai-employment.org/adonaimetteam7.html	4
http://adonai-employment.org/adonaisuccess7.html	5
http://adonai-employment.org/contact_uslocation	6
http://adonai-employment.org/employer_services	7
http://adonai-employment.org/home	8
http://adonai-employment.org/index.html	9

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

http://adonai-employment.org/job_seeker_services	10
http://adonai-employment.org/meet_our_team	11
http://adonai-employment.org/meet_our_team/	12
http://adonai-employment.org/resourceslinks	13
http://adonai-employment.org/robots.txt	14
http://adonai-employment.org/success_stories	15
http://search.rr.com/?q=Adonai+Employment%2C+Milwaukee&cat=web	16
http://search.yahoo.com/search;_ylt=Ak7PHpHUrPuObbkqBCK3p0CbvZx4?fr=yfp-t-701-s&toggle=1&cop=mss&ei=UTF8&p=adonai%20employment%20milwaukee	17
http://search.yahoo.com/search;_ylt=An7LkiOmQWAeOmB1kYFqxsimN3wV?p=ado	18
http://search.yahoo.com/search?ei=UTF-8&fr=ytf-&p=jobs+with+adonai	19
http://trafficfaker.com	20
http://www.adonai-employment.org	21
http://www.adonai-employment.org/	22
http://www.adonai-employment.org/about_us	23
http://www.adonai-employment.org/contact_uslocation	24
http://www.adonai-employment.org/home	25
http://www.adonai-employment.org/meet_our_team	26
http://www.adonai-employment.org/resourceslinks	27
http://www.bing.com/search?q=adonai+wisconsin&src=IE-SearchBox&Form=IE8SRC	28
http://www.exalead.com/	29
http://www.google.co.uk/url?sa=t&source=web&cd=4&ved=0CDQQFjAD&url=http%3A%2F%2Fadonaiemployment.org%2F&rct=j&q=adonai%20nursing%20agency%20&ei=oY9WTbbtKoSr8QOB7fDADQ&usg=AFQjCNHbhxbDkZOJLB4vxoM3Ltb57Hv-Lw	30
http://www.google.co.za/url?sa=t&source=web&cd=6&ved=0CDcQFjAF&url=http%3A%2F%2Fadonaiemployment.org%2F&rct=j&q=adonai%20recruitment&ei=mpZkTbyCDsyFhQeXjYm-Bw&usg=AFQjCNHbhxbDkZOJLB4vxoM3Ltb57Hv-Lw	31
http://www.google.co.za/url?sa=t&source=web&cd=7&ved=0CEQQFjAG&url=http%3A%2F%2Fadonaiemployment.org%2F&rct=j&q=adonai%20recruitment&ei=jUJSTaKJH4m0hAfb0cW2CQ&usg=AFQjCNHbhxbDkZOJLB4vxoM3Ltb57HvLw&sig2=Wyf5Cn7c0G6RVtiHgJ86kQ	32
HTTP://WWW.GOOGLE.COM/SEARCH?CLIENT=SAFARI&RLS=EN&Q=ADONAI+EMPLOYMENT+MILWAUKEE&IE=UTF-8&OE=UTF-8	33
http://www.google.com/search?hl=en&q=adonai+employment+milwaukee&aq=1v&aqi=g1g-v5&aql=&oq=adonai+employment+	34
http://www.google.com/search?hl=en&rlz=1T4RNTN_enUS337US338&q=Adonai+el	35
http://www.google.com/search?hl=en&source=hp&q=adonai+employment&aq=f&aqi=g1g-v6&aql=&oq=	36
http://www.google.com/search?hl=en&source=hp&q=adonai+employment+inc&aq=0&aqi=g1g-v6&aql=&oq=adonai+empl	37
http://www.google.com/search?q=adonai+employment&ie=utf-8&oe=utf8&aq=t&rls=org.mozilla:en-US:official&client=firefox-a	38
http://www.google.com/search?q=adonai+employment&rls=com.microsoft	39
http://www.google.com/url?sa=t&source=web&cd=1&sqi=2&ved=0CBcQFjAA&url=http%3A%2F%2Fadonaiemployment.org%2F&rct=j&q=adonai%20employment&ei=XRUVtby3Eoe2tweRqsiRCg&usg=AFQjCNHbhxbDkZOJLB4vxoM3Ltb57Hv-Lw	40

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Figura 12. Log Parametrizado Adonai Employment.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
1	IP	FECHA	HORA	MINUTO	SEGUNDO	ESTADO	BYTE	URL	NAVEGADOR							
2	A		0	13	40	24	1	0	2							
3	A		0	13	42	15	1	0	10							
4	A		0	14	14	37	1	0	22							
5	A		0	14	14	39	1	0	23							
6	A		0	14	14	40	1	1	26							
7	A		0	14	14	40	1	1	26							
8	A		0	14	14	40	1	1	26							
9	A		0	14	14	40	1	3	26							
10	A		0	14	14	40	1	1	26							
11	A		0	14	14	40	1	3	26							
12	A		0	14	14	40	1	1	26							
13	A		0	14	16	37	1	0	26							
14	A		0	14	16	53	1	0	25							
15	A		0	14	30	7	1	0	26							
16	A		0	18	4	20	1	0	2							
17	A		0	18	4	21	1	3	11							
18	A		0	18	4	21	1	1	11							
19	A		0	18	4	21	1	1	11							
20	A		0	18	4	24	1	1	11							
21	C		1	2	28	35	1	0	29							
22	C		1	2	28	36	1	1	29							
23	A		1	17	54	33	1	0	49							

Se convirtió el archivo parametrizado a un documento con extensión (.Arff)

Figura 13. Documento Extensión .Arff.

```

Adonai: Bloc de notas
Archivo Edición Formato Ver Ayuda
Relation adonai
@attribute IP {A, B, C}
@attribute FECHA {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19}
@attribute HORA {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23}
@attribute MINUTO {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 3}
@attribute SEGUNDO {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35}
@attribute ESTADO {1}
@attribute BYTE {0, 1, 2, 3}
@attribute URL {0, 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31, 32, 33, 34, 35, 36}
@attribute NAVEGADOR {0, 1, 2, 3}

@data
A,0,13,40,24,1,0,2,1
A,0,13,42,15,1,0,10,1
A,0,14,14,37,1,0,22,0
A,0,14,14,39,1,0,23,0
A,0,14,14,40,1,1,26,0
A,0,14,14,40,1,1,26,0
A,0,14,14,40,1,3,26,0
A,0,14,14,40,1,1,26,0
A,0,14,14,40,1,1,26,0
A,0,14,16,37,1,0,26,0
A,0,14,16,53,1,0,25,0
A,0,14,30,7,1,0,26,0
A,0,18,4,20,1,0,2,1
A,0,18,4,21,1,3,11,1
A,0,18,4,21,1,1,11,1
A,0,18,4,21,1,1,11,1
A,0,18,4,24,1,1,11,1
C,1,2,28,35,1,0,29,1
C,1,2,28,36,1,1,29,1
A,1,17,54,33,1,0,49,1
A,1,17,54,42,1,1,11,1
A,1,17,54,42,1,1,11,1
A,1,17,54,42,1,1,11,1
A,1,17,54,42,1,1,11,1
A,1,17,54,43,1,3,11,1
A,1,17,54,43,1,1,11,1
A,1,18,21,30,1,0,11,1
A,1,18,39,37,1,0,2,1
A,1,18,39,37,1,1,11,1
A,1,18,39,38,1,1,11,1
A,1,18,39,38,1,1,11,1
A,1,22,11,3,1,1,11,1
A,1,22,11,3,1,1,11,1
A,1,22,11,3,1,1,11,1
A,1,22,11,3,1,1,11,1

```

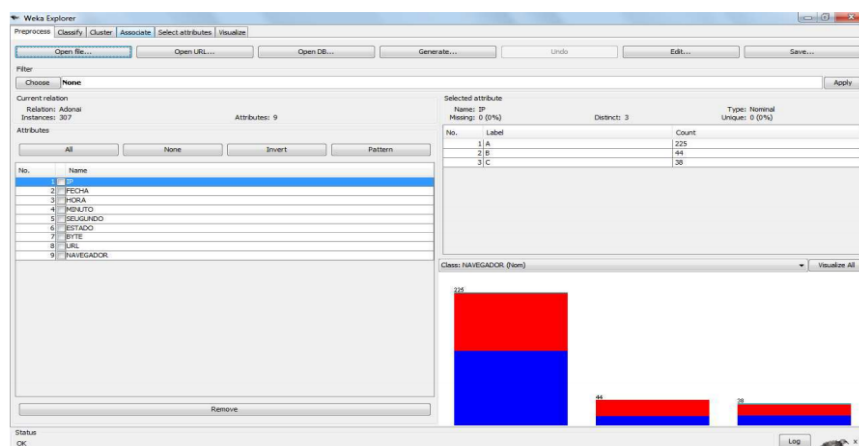
PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

10. Análisis en el Weka

Después de parametrizar el log por medio de la herramienta Excel, se realizó un archivo con formato .Arff, el cual fue interpretado por el software libre Weka para aplicar el algoritmo A priori, que generó unas reglas de asociación para predecir el comportamiento de los usuarios.

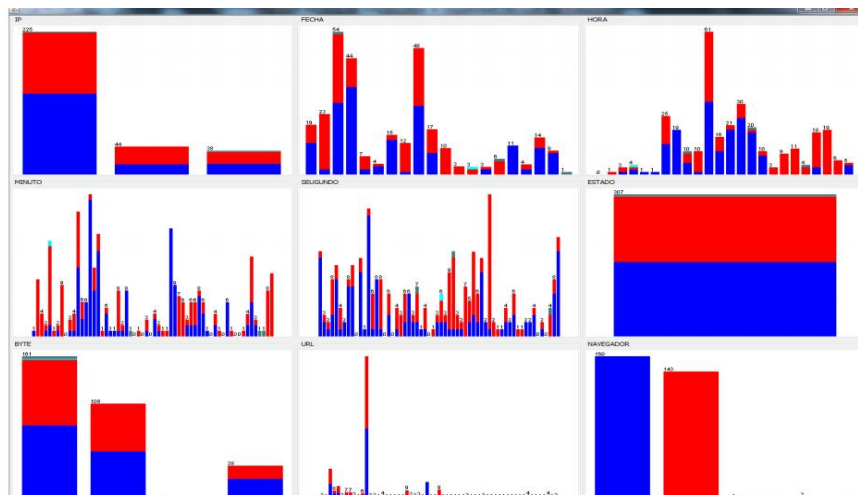
Al abrir el archivo .Arff se puede notar que contiene 9 atributos y 307 instancias, además, se observa como primera medida la gráfica correspondiente a la columna IP.

Figura 14. *Gráfica Columna IP.*



De la misma manera se presentan cada una de las gráficas de los 9 atributos: IP, Fecha, Hora, Minuto, Segundo, Estado, Bytes, URL y Navegador.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Figura 15. Gráfica de los 9 Atributos.

Sin embargo, el enfoque estaba en extraer los patrones de comportamiento, razón por la cual se sacaron las Reglas de Asociación a priori predictivo. Cabe resaltar que el “Estado” corresponde a si la petición fue exitosa, es decir, si el usuario pudo acceder al dominio adonai-employment.org o no, donde los valores son específicamente 1 y 2, siendo el primero equivalente a “exitosa” y el segundo a “no lograda”.

El proceso comenzó con un conjunto de datos procedentes de la interacción del usuario con el sitio el cual alojó toda esta información en el log del servidor de la empresa, a continuación se presentan las reglas.

11. Reglas de Asociación Generadas por el Algoritmo A Priori en Weka

1. IP=A 225 ==> ESTADO=1 225 conf:(1)
2. BYTE=0 161 ==> ESTADO=1 161 conf:(1)
3. NAVEGADOR=0 160 ==> ESTADO=1 160 conf:(1)

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

4. NAVEGADOR=1 143 ==> ESTADO=1 143 conf:(1)
5. URL=11 133 ==> ESTADO=1 133 conf:(1)
6. IP=A NAVEGADOR=0 127 ==> ESTADO=1 127 conf:(1)
7. IP=A BYTE=0 110 ==> ESTADO=1 110 conf:(1)
8. BYTE=1 108 ==> ESTADO=1 108 conf:(1)
9. IP=A URL=11 101 ==> ESTADO=1 101 conf:(1)
10. IP=A NAVEGADOR=1 96 ==> ESTADO=1 96 conf:(1)

11.1 Análisis de las Reglas de Asociación

Al aplicar las reglas de asociación se pudo determinar lo siguiente:

1. A partir de la primera regla se pudo inferir con un grado de confianza del 100% que los 255 usuarios con IP de clase A tuvieron una petición exitosa al acceder a la página www.adonai-employment.org.
2. Con una confianza del 100% se pudo observar que 161 usuarios tuvieron una transferencia de bytes en el rango de 0 a 50.000 y a su vez, sus peticiones fueron exitosas.
3. Se pudo observar con un grado de confianza del 100% que 160 usuarios entraron al sitio web de Adonai Employment y tuvieron peticiones exitosas desde el navegador Mozilla Firefox.
4. El navegador Chrome 86 fue utilizado en 143 consultas de las cuales todas fueron exitosas ya que consiguieron entrar al sitio web, lo que generó un acierto del 100% en la ejecución de las consultas.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

5. 133 peticiones exitosas fueron realizadas al acceder a la página web a través del link “http://adonai-employment.org/meet_our_team” anotándose un acierto del 100%.

6. Con el navegador Mozilla Firefox 127 usuarios con IP de clase A realizaron 127 consultas exitosas lo cual creó confianza en el navegador al momento de realizar las peticiones al servidor.

7. 108 usuarios con una IP de tipo A tuvieron una transferencia de bytes en el rango de 0 a 50.000 y su petición fue exitosa en el 100% de los casos.

8. En 108 peticiones realizadas por los usuarios de la página de Adonai Employment ocurrió una transferencia de bytes en el rango de 0 a 50.000 con un grado de confianza del 100%.

9. 101 usuarios con un tipo de IP A tuvieron peticiones exitosas por medio del URL 11, es decir, a través de “http://adonai-employment.org/meet_our_team”, lo que evidenció un acierto del 100% al ingresar a través de esta URL.

10. El navegador Chrome 86 fue utilizado por 96 usuarios con un tipo de IP A los cuales tuvieron 96 peticiones exitosas lo cual proporcionó un acierto del 100% en las consultas realizadas a través de este navegador.

12. Conclusiones

Los resultados más representativos obtenidos mediante la aplicación del software libre Analog se sintetizan a continuación:

- Los días martes, miércoles y jueves (en orden descendente) se presentó mayor cantidad de peticiones respecto a los demás días de la semana en la página web de la empresa Adonai Employment. El flujo de estos tres días representa exactamente el 52,32% del total de las peticiones realizadas en el periodo del 06 al 27 de febrero del 2020.
- A las 10 am se generó el mayor número de peticiones en comparación con las demás horas del día, específicamente 76. Además, se pudo deducir que los usuarios de la empresa Adonai Employment ingresan a la página web principalmente en horas de la mañana, entre las 6 y 11 am, ya que el 34,01% de las 732 peticiones exitosas coinciden con este horario. Por su parte, en segundo lugar, la mayoría de consultas son realizadas entre las 12 y 17 horas con un 24,25%. Lo que indica que más de la mitad (58,26%) de los usuarios de la empresa Adonai Employment accedieron entre las 6 am y las 5 pm al dominio adonai-employment.org en el periodo que comprende desde el 06 hasta el 27 de febrero del 2020.
- La subred o familias de IP que accedieron con mayor frecuencia al servidor de Adonai Employment fue la 207.46. con un total de 96 peticiones realizadas durante las 3 semanas analizadas (06-27 de febrero 2020). A su vez, se estimó que las 10 primeras organizaciones de las 75 identificadas y ordenadas por número de peticiones comprendieron el 54,37% del total de las peticiones efectuadas en este periodo. Es decir, que el 13,33% de las organizaciones realizó más del 50% de las consultas en la página de la empresa Adonai Employment.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

- El sistema operativo más utilizado por los usuarios de adonai-employment.org es Windows al comprobar que 311 peticiones se realizan a través del mismo, es decir, el 42,49% de las peticiones exitosas fueron ejecutadas por medio del sistema operativo Windows.
- Los usuarios de adonai-employment.org solicitan recurso “pesados” ya que en el 76,37% de las peticiones llevadas a cabo se identificó que el tamaño requerido de los archivos osciló entre los 10kB y 100kB.

A partir de lo mencionado anteriormente, fue posible observar que los usuarios de la página web de la empresa Adonai Employment conformados en parte por las personas del área metropolitana de Milwaukee con cierta desventaja o discapacidad, presentan características comunes en sus visitas de navegación en busca de empleo, como lo son los días de la semana y la hora en la que realizan sus peticiones o consultas, ya que principalmente comprenden la mañana (entre las 6 y 11 am) durante tres días de la semana: martes, miércoles y jueves; esta información es pertinente para ser analizada directamente por la organización Adonai Employment con el fin de implementar mejoras en el diseño de su página web, como realizar las publicaciones que impliquen información sobre ofertas laborales los fines de semana para lograr abarcar más del 50% de los usuarios y dar respuesta a sus necesidades, lo que significa cumplir con el objetivo de brindar apoyo a través de un empleo a la población vulnerable en la ciudad propuesto por Adonai Employment.

Por otro lado, con relación a los perfiles y características de los usuarios basados en las reglas de asociación, estos presentaron el siguiente comportamiento:

- Los usuarios de adonai-employment.org que accedieron en el periodo del 06 al 27 de febrero del 2020 a <http://www.adonai-employment.org> lo hicieron a través de dos navegadores principalmente: Mozilla Firefox y Chrome 86 y de la misma forma, lograron

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

búsquedas exitosas en todas las oportunidades, ya que como se expresó anteriormente, la confianza obtenida en las reglas de asociación desarrolladas en el software Weka fue de 1, que corresponde al 100%.

- El link de acceso “http://adonai-employment.org/meet_our_team” se considera confiable, ya que las 133 peticiones realizadas a través del mismo fueron exitosas en su totalidad, correspondientes al 18,17% de las 732 peticiones realizadas del 06 al 27 de febrero del 2020.
- Se evidenció una confianza del 100% de la misma manera en el link “http://adonai-employment.org/meet_our_team”, ya que los 101 usuarios que lo utilizaron lograron realizar sus peticiones con éxito en el 100% de los casos.

De acuerdo a lo mencionado anteriormente correspondiente al análisis de la información suministrada a través de las reglas de asociación, se pudo determinar que las respuestas del servidor de la página web de la empresa Adonai Employment son satisfactorias para los usuarios, independientemente del navegador y el sistema operativo utilizado por los mismos. Cuando se habla de confianza se hace referencia a la probabilidad existente de que los usuarios o el grupo de usuarios que cumplen con una característica o atributo X, como lo es el acceso a adonai-employment a través del navegador Mozilla Firefox, también tengan en común el atributo Y, como por ejemplo el estado exitoso de la petición. Es decir, “si X entonces Y”.

Es así, como se garantiza que las tendencias de los usuarios de la organización de acceder a www.adonai-employment.org a través de los navegadores Mozilla Firefox y Chrome 86 generan peticiones satisfactorias para los usuarios, ya que “si se utiliza el navegador Mozilla Firefox o Chrome 86 entonces la petición es exitosa” de acuerdo a las reglas de asociación.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

Por su parte, los proyectos de web mining permiten a las empresas saber lo que realmente está pasando en su sitio web porque pueden conocer las características principales y los perfiles de sus visitantes, además, de los productos y servicios que sus clientes compran o solicitan, así como saber lo que les gustaría encontrar a través de sus búsquedas. Estos y otros interrogantes se pueden responder a partir de los datos de una empresa u organización y su sitio web.

Por consiguiente, se emplearon técnicas de data mining según el objetivo de la empresa Adonai Employment, que fue principalmente segmentar a sus usuarios y describir el comportamiento de estos para posteriormente realizar el análisis del mercado visto desde un nuevo contexto, es decir, desde el uso de internet y el almacén de grandes volúmenes de datos en su funcionamiento diario.

En definitiva, por medio del descubrimiento de conocimientos en bases de datos, en este caso utilizando el log del servidor de Adonai Employment, se pudo identificar patrones útiles para la organización o empresa identificados anteriormente, a partir de los datos de navegación dejados en el log del servidor de una manera automática utilizando grandes cantidades de datos. Por ello, se considera que el desarrollo de estos proyectos de investigación y la minería de datos es vital para la extracción de conocimiento implícito en grandes bases de datos y tiene un papel fundamental en el proceso de convertir en explícito al conocimiento implícito y en las distintas etapas del proceso de gestión del conocimiento en una empresa u organización.

13. Recomendaciones

Se recomienda a la empresa Adonai Employment hacer uso de la información obtenida a partir de las técnicas de minería de datos para identificar posibles falencias con relación a peticiones que no pudieron ser ejecutadas por parte de los usuarios debido a un acceso inválido o

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

problemas directamente relacionados con el servidor, así como para evitar la sobrecarga de su servidor web, lo cual se logra inferir a través de los softwares utilizados previamente. De la misma manera, implementar estrategias que permitan a los usuarios encontrar la información que necesitan en el momento que la requieren, para ello se pueden basar en los comportamientos de búsqueda analizados previamente.

Se sugiere, además, continuar con la implementación de proyectos enfocados en el web mining y data mining debido a la seguridad que existe tanto para la empresa Adonai Employment como para cualquier otra, porque permite la detección de accesos inusuales a datos privados, la implementación de equipos en una empresa según los requerimientos técnicos y la distribución de datos con el fin de manejar eficientemente el tráfico de un sitio, de igual forma, permite el rediseño de sitios web y todos estos resultados, sirven de soporte a decisiones de marketing o negocios por medio de la determinación de conductas o rasgos de los clientes que realizan ciertas acciones según la naturaleza de la empresa.

En resumen, el uso de las técnicas en minería de datos permite a las empresas conocer lo que realmente está pasando en su sitio web, las tendencias principales y características de sus visitantes, como también conocer por qué ciertos clientes compran o solicitan un determinado producto y para ello se desarrollan estudios y modelos utilizando metodologías y técnicas de data mining, gracias a que su sitio web genera un funcionamiento diario y la información está contenida en el log del servidor de la empresa.

Referencias

- [1] C. C. d. ONG, «ccong.org,» 08 2016. [En línea]. Available: https://ccong.org.co/files/728_at_Lo%20que%20hay%20que%20saber%20de%20las%20ESAL,%20agosto%20de%202016.pdf.
- [2] A. E. Inc, «Adonai Employment,» 2008. [En línea]. Available: <http://www.adonai-employment.org/>.
- [3] C. L. Pineda Tarazona, «Repositorio Institucional USTA,» 2018. [En línea]. Available: <https://repository.usta.edu.co/bitstream/handle/11634/12665/2018cristianpineda.pdf?sequence=1&isAllowed=y>.
- [4] V. A. Sanabria Ruiz, «Tangara Uis,» 2017. [En línea]. Available: <http://tangara.uis.edu.co/biblioweb/tesis/2017/168596.pdf>.
- [5] E. A. Rojas Gutiérrez y J. S. Aguilar, «Repositorio U Católica,» 2017. [En línea]. Available: <https://repository.ucatolica.edu.co/bitstream/10983/15329/1/Trabajo%20de%20grado.pdf>. [Último acceso: 23 05 2020].
- [6] M. García, «EcuRed,» [En línea]. Available: <https://www.ecured.cu/Weka>. [Último acceso: 23 05 2020].
- [7] Ó. Buenaposada Cano, «Openaccess UOC,» 15 01 2017. [En línea]. Available: <http://openaccess.uoc.edu/webapps/o2/bitstream/10609/59326/6/obuenaposadaTFG0117mem%C3%B2ria.pdf>. [Último acceso: 23 05 2020].
- [8] D. Gutiérrez Cruz, Y. A. Albarrá Fernández y R. Rico Molina, «Uaemex,» 10 2017. [En línea]. Available: http://ri.uaemex.mx/bitstream/handle/20.500.11799/69982/secme-30553_1.pdf?sequence=1&isAllowed=y. [Último acceso: 23 05 2020].
- [9] E. A. Zamora Rodríguez, «Biblioteca PUCV,» 07 2008. [En línea]. Available: http://opac.pucv.cl/pucv_txt/txt-9000/UCI9016_01.pdf. [Último acceso: 23 05 2020].
- [10] J. E. Velasco López, «E-Prints Complutense,» 2013. [En línea]. Available: <https://eprints.ucm.es/25812/1/Trabajo%20Fin%20Master%20Jorge%20Velasco%20%2081%29.pdf>. [Último acceso: 23 05 2020].
- [11] M. A. North, «Biblioteca Virtual FAHUSAC,» 01 10 2012. [En línea]. Available: <https://bvhumanidades.usac.edu.gt/items/show/875>. [Último acceso: 23 05 2020].

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

- [12] M. Á. Grández Márquez, «Repositorio Usil,» 2017. [En línea]. Available: http://repositorio.usil.edu.pe/bitstream/USIL/2763/1/2017_Granda_Aplicacion-de-mineria-datos.pdf.
- [13] E. M. Ruiz Lobaina y P. L. Romero Suárez , «Scielo,» 26 07 2018. [En línea]. Available: 2018.
- [14] Á. Rodríguez H, «Revista Consultoría,» 09 09 2019. [En línea]. Available: <https://revistaconsultoria.com.mx/aplicacion-la-mineria-datos-recursos-humanos/>.
- [15] F. C. Noreña, «Repositorio Universidad EIA,» 2015. [En línea]. Available: https://repository.eia.edu.co/bitstream/11190/2288/1/CaladFelipe_2015_SegmentacionClientesAutomatizada.pdf.
- [16] L. C. Caro Martínez, «Repositorio Los Libertadores,» 2016. [En línea]. Available: <https://repository.libertadores.edu.co/bitstream/handle/11371/727/CaroMart%C3%ADnezLauraConsuelo.pdf?sequence=2&isAllowed=y>.
- [17] A. I. Oviedo Carrascal y J. Jiménez Giraldo, «Minería de datos educativos: Análisis del desempeño de estudiantes de ingeniería en las pruebas Saber Pro,» *Revista Politécnica ISSN*, vol. 15, n° 29, p. 13, 2019.
- [18] C. A. Ortiz Cárdenas, «CRAI USTA,» 2015. [En línea]. Available: <http://crai.ustabuca.edu.co/>.
- [19] A. Bustamante, E. Galvis y L. C. Gómez, «Perfil de la investigación sobre inteligencia de negocios,» *UIS Ingenierías*, vol. 15, n° 1, p. 11, 2016.
- [20] L. Y. Caicedo Chacón, C. N. Cárdenas Parra, J. S. Müller Rueda y J. T. Ortiz Bernal , «Aplicación para la gestión y el análisis de información relacionada con la deserción estudiantil universitaria,» *RCC - Revista Colombiana de Computación*, vol. 20, n° 2, p. 14, 2019.
- [21] R. Hernández Sampieri, *Metodología de la Investigación*, México D.F.: McGraw Hill, 2014.
- [22] J. Hurtado de Barrera , *Metodología de la Investigación Holística*, Caracas: Sygal - Servicios y Proyecciones para América Latina, 2000.
- [23] Á. Pérez Niño, «ResearchGate,» 2015. [En línea]. Available: https://www.researchgate.net/publication/321173529_Web_usage_mining_para_la_identificacion_de_patrones_de_comportamiento_de_usuarios_mediante_el_uso_de_herramientas_tecnologicas.

PATRONES DE COMPORTAMIENTO ADONAI EMPLOYMENT

- [24] C. A. R. Sarmiento, «Biblioteca Virtual UIS,» 15 11 2012. [En línea]. Available: <http://tangara.uis.edu.co/biblioweb/tesis/2012/146524.pdf>.