

Aplicación del modelo matemático similitud de coseno para identificar gustos de usuario en una red social

Application of the mathematical model similarity of cosine, to identify user preferences in a social network.

Carlos Augusto Osorio Vélez¹
carlos.osorio07@usc.edu.co

Christian Sánchez Quiroga¹
christian.sanchez01@usc.edu.co

Ciro Antonio Dussan Clavijo²

Universidad Santiago de Cali, Facultad de Ingeniería, Programa de Ingeniería de sistemas¹.
Ingeniero de sistemas, MSc , Profesor titular, Facultad de Ingeniería, Universidad Santiago de Cali².

Resumen

La presente investigación establece como objetivo general la aplicación del modelo matemático similitud de coseno para identificar gustos de usuario en una red social, incorporando un algoritmo que proporciona precisión, cohesión de datos y rendimiento. La metodología utilizada es aplicada-cuantitativa, enfocada en las necesidades y búsquedas prácticas para la ejecución de procesos operacionales y de sistemas, con el modelo Scrum se desarrolló el marco de trabajo que planifico los Sprints con fases semanales durante los meses de noviembre y abril, realizando las debidas inspecciones del modelo. Los resultados destacan que la arquitectura del software para la aplicación del algoritmo similitud de coseno en la red social, constituida en el lenguaje de programación, framework, base de datos por medio de Python, Flask y MySQL ha propuesto una estructura de funcionalidad tecnológica con una serie de componentes hacia un fin específico, por cuanto el desarrollo del aplicativo promueve un producto de utilidad, al ejecutar el algoritmo se logró un conteo exacto de los datos del 100% al 25%, teniendo en cuenta la afinidad de las preferencias establecidas por el cliente en la búsqueda, la integración del algoritmo de coseno al sistema web con el modelo de red social, permitió recomendar sitios de interés teniendo en cuenta la información suministrada, dando un uso de tipo aplicado a tendencias actuales como son las redes sociales y los algoritmos basado en sistemas de recomendación que en el presente caso utilizo el filtrado basado en contenido, reconociendo características y atributos de los elementos de búsqueda, para sugerir productos o servicios que le pueden interesar al cliente.

Palabras Claves: Algoritmo, Gustos, Red social, Similitud de coseno, Usuario.

Abstract

The present research establishes as a general objective the application of the mathematical model similarity of cosine to identify user tastes in a social network, incorporating an algorithm that provides precision, data cohesion and performance. The methodology used is applied-quantitative, focused on the needs and practical searches for the execution of operational processes and systems, with the Scrum model the framework was developed that planned the Sprints with weekly phases during the months of November and April, performing the proper model inspections. The results highlight that the software architecture for the application of the cosine similarity algorithm in the social network, constituted in the programming language, framework, database through Python, Flask and MySQL, has proposed a structure of technological functionality with a series of components towards a specific end, since the development of the application promotes a useful product, when executing the algorithm an exact count of the data from 100% to 25% was achieved, taking into account the affinity of the preferences established by the client in the search, the integration of the cosine algorithm to the web system with the social network model, allowed to recommend sites of interest taking into account the information provided, using a type applied to current trends such as social networks and algorithms based on recommendation systems that in the present case I use content-based filtering recognizing characteristics and attributes Search items to suggest products or services that the customer may be interested in.

Keywords: Algorithm, Likes, Social network, Cosine similarity, User.

1. INTRODUCCIÓN

Laudon (2018) menciona que los consumidores al momento de utilizar portales electrónicos de compra, servicios, redes sociales, transmisión digital y sitios de entretenimiento requieren que los contenidos de interés estén disponibles y puedan accederse de manera simple, en tal escenario, el reconocimiento de las preferencias del cliente es posible, si los sitios web incorporan sistemas de recomendación para el análisis de los gustos del usuario. Para tal fin, Hoi y Tieu (2018) reconocen que el desarrollo de algoritmos constituye un proceso sistemático para las organizaciones, puesto que facilita a los clientes de los sitios de comercio apropiarse de información fundamental para la toma de decisiones. Según Bayona y Pardo (2015) el algoritmo presenta un método con una serie de instrucciones, que permiten resolver un determinado problema, fundamentales en el campo de la programación (p.15). El desarrollo de algoritmos es una aplicación práctica de la ingeniería, para el conocimiento científico de diseño y construcción de programas de cómputo, se establece, como parte del proceso de innovación tecnológica, en miras de aportar soluciones a diferentes necesidades empresariales (Saha y Ghosh, 2020).

En años recientes, se destaca el empleo de algoritmos de similitud en empresas de base tecnológica, como Netflix, Amazon, YouTube y Facebook, donde la optimización comparado con métricas antiguas ha demostrado que las implementaciones con algoritmos presentan beneficios para usuarios y empresas (Ibrahim, Husnu y Kuo-Chi, 2020, pág. 113). Las ventajas del algoritmo de similitud del coseno corresponden a precisión, cohesión de datos y generación de resultados (Belazzoug y Brahimi, 2020). Es así, como desde la ingeniería se diseñan algoritmos de similitud, que ejecutan el cálculo matemático para determinar semejanzas que existen entre un colectivo de datos. No obstante, la efectividad del algoritmo depende del dominio para el cual están diseñados, priorizando un proceso que brinde precisión en la recomendación hacia el consumidor (Texier y Guisti, 2019).

Generalmente el usuario interactúa en diferentes sitios web, buscando opciones de consumo, por tanto la sugerencia del producto o servicio derivada del algoritmo debe estar acorde con su perfil y características de preferencia, en ese panorama, Mendoza y Pérez (2019) indican que si el algoritmo no sugiere un ítem interesante, la consecuencia será que no se genere una venta en el caso de una tienda online, o en el caso de una red social que no se interactúe con el contenido deseado por el usuario, realizando recomendaciones erradas según las preferencias y gustos del comprador, debido a información innecesaria y poco precisa. En ese sentido, el diseño del modelo matemático debe considerar la incorporación de un algoritmo que proporcione confiabilidad, rendimiento y porcentajes de no fallo (Tang, 2019). La estructuración del algoritmo requiere de conocimientos previos y creatividad para una implementación efectiva en la resolución del problema. Ya que el algoritmo, en el campo de la ingeniería se establece como una secuencia de instrucciones que representan un modelo de solución (Jabba y Mejía, 2017).

Diversos estudios desde el campo de la investigación han motivado el desarrollo de algoritmos de similitud en el contexto empresarial, a fin de mejorar procesos operativos y de relación con el usuario, en ese panorama González (2016) diseña un algoritmo de similitud del coseno con filtrado colaborativo, en una plataforma compuesta por empresas que se dedican a la distribución y venta de comestibles. Los hallazgos indican que una vez se han ajustado los parámetros necesarios para el funcionamiento del algoritmo, se logran resultados óptimos en términos de validación y rendimiento, destacando que la lista de recomendación es mejor a medida que se adicionan más elementos, puesto que es más fácil para el usuario encontrar elementos relevantes entre más ítems recomendados se integren al modelo del algoritmo, por ende, con la recomendación de los productos se consigue mayor grado de satisfacción del consumidor e información valiosa para la empresa.

Lozano y Fuentes (2017) demuestran la utilidad que ofrece un software desarrollado con algoritmo de similitud de coseno para una entidad del sector bancario, en ese propósito se emplean contenidos que interesan al usuario relacionados con consultas bancarias, productos financieros y experiencias del cliente frente a la atención y servicio, para mediante un estudio de las similitudes arrojadas por el algoritmo, proceder a la comparación entre la consulta formulada por el usuario y las categorías identificadas, los resultados indican que la aplicación de esta metodología genera ahorro de tiempo de búsqueda y recomendaciones sobre problemas financieros del que tratan las consultas, la información que le proporciona el sistema al cliente permite mejorar la negociación y relación con el banco, al incrementar conocimientos financieros y organizacionales con la entidad.

Ghobar (2019) desarrolla un software con algoritmos de similitud, sugiriendo productos y servicios en una tienda Online, en relación con los gustos, perfil o ítems consultados por el usuario, la implementación del algoritmo reconoce patrones de entrada por parte del usuario, indicando el número de grupos que se desea crear. La construcción del programa utiliza filtrado colaborativo asociado al contenido de la tienda; demostrando ser eficiente para la generación de recomendaciones, en términos de las características del usuario como gustos y productos frecuentes de compra. El programa ha utilizado el Model-based techniques en cascada, compuesto por tareas de agrupado y asociación, los resultados demuestran que el modelo tiene capacidad de convertirse en un sistema de recomendaciones económico y eficiente, en aquellos campos en los que se recogen pocos datos de usuario.

Harbir et al. (2020) estructuran un sistema de programación utilizando el algoritmo de similitud de coseno para una plataforma de contenido digital, mediante este modelo se ofrecen recomendaciones a los usuarios sobre películas, series y detecta problemas comunes a la hora de acceder a la plataforma, teniendo en cuenta las calificaciones que otorgan los usuarios (p.22). El sistema establece el filtrado basado en contenido, creando perfiles en función de gustos y preferencias del cliente, utilizando la información recopilada para la recomendación del producto. La fórmula matemática midió los gustos de los usuarios con las películas basado en similitudes de diferentes propiedades, matemáticamente mostro el coseno del ángulo de dos vectores, proyectado de manera multidimensional, por ende, el programa recomendó las mejores películas al usuario según la similitud del coseno, este modelo genera eficiencia y productividad a la plataforma.

Teniendo en cuenta lo anterior, la presente investigación establece como objetivo general demostrar la aplicación del algoritmo similitud de coseno para identificar gustos de usuario en una red social, enfocado en el sector de los restaurantes y sitios de entretenimiento nocturno en Santiago de Cali. Frente a los aspectos que destacan el modelo matemático de similitud de coseno, se identifica que los vectores cumplen con la condición de ser similares si son paralelos, y en el caso que sean diferentes ser ortogonales o perpendiculares, posteriormente a los vectores se les da un valor que corresponda con la dimensión en igual cantidad de veces, de esta manera se puede detectar cuantas veces aparece un término o un valor (Hoda y Salleh, 2018). Es importante indicar que la característica principal de este tipo de algoritmo es la búsqueda y recuperación de la información, actualmente usada en la minería de textos. La minería de textos es un campo de investigación del procesamiento automático de datos, que surge con la necesidad de desarrollar algoritmos capaces de procesar información estructurada y textual (Belsky, 2019). Así, al ejecutar el algoritmo se logra un conteo exacto de los datos, la integración del algoritmo de coseno en un sistema web con el modelo de red social, le permitirá acceder y administrar la información suministrada por los usuarios en la red, frente los gustos, preferencias y tendencias de consumo de productos y servicios que ofrecen los distintos establecimientos nocturnos y restaurantes de Santiago de Cali.

Según Global Nielsen (2018) con datos de la asociación de restaurantes nocturnos en Colombia, los consumidores a la hora de interactuar con redes sociales buscan sugerencias en el tipo de comida u opción gastronómica, lugar de entretenimiento, precio del producto y horario de funcionamiento del sitio. Por tanto, se propone que el algoritmo recomiende restaurantes y sitios nocturnos de la ciudad, basado en los gustos del usuario, estableciendo la sugerencia del lugar con un grado de precisión exacto, basado en las preferencias del cliente, específicamente en aspectos como el tipo de comida (vegetariana, callejera, italiana, colombiana), lugar de entretenimiento (bar, discoteca, restaurante), ambiente (juvenil, clásico, familiar) y música (rock, salsa, bachata, reggaetón). En consecuencia, la estructura del artículo se puntualiza en cuatro secciones: metodología y tres objetivos específicos que presentan (1) características y funcionalidad de la red social, (2) arquitectura del software para la aplicación del algoritmo similitud de coseno en la red social y (3) funcionamiento del modelo matemático de similitud de coseno para reconocer gustos de usuario en una red social.

2. METODOLOGÍA

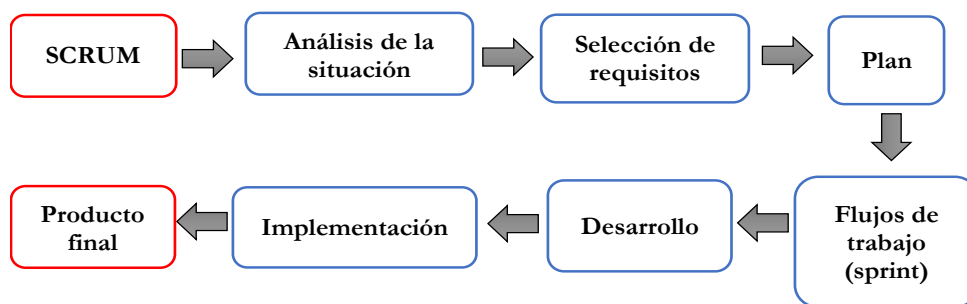
2.1 Tipo de estudio

El tipo de estudio para el desarrollo de la investigación es aplicado-cuantitativo, enfocado en las necesidades y búsquedas prácticas para la ejecución de procesos operacionales y de sistemas, en ese aspecto, Lozada (2018) señala que la investigación aplicada se fundamenta en hallazgos tecnológicos de investigación, generando conocimientos a partir de procedimientos que involucran elementos teóricos y prácticos del producto implementado. Es importante destacar que la investigación aplicada, se apoya en lo cuantitativo por tener un diseño estructurado y secuenciado que se fundamenta en el cálculo matemático, generando datos estandarizados y probatorios que se derivan del análisis del modelo matemático utilizado por el investigador, en el presente caso se evidencia en la aplicación del algoritmo de similitud del coseno en la red social.

2.2 Método

El método corresponde al modelo Scrum, proporcionando las tareas para el diseño del lenguaje de programación, estructura secuencial, framework y procesamiento de datos para la aplicación del algoritmo similitud de coseno en una red social basado en gustos de usuario, cabe mencionar que Scrum establece un marco de trabajo, que controla y planifica el proceso de construcción del sistema. Urteaga (2018) menciona que Scrum es una metodología que permite seguir de forma clara el avance de las tareas a realizar, siendo apropiada para desarrollar proyectos que requieren agilidad y flexibilidad con buenas prácticas, especialmente la ejecución de modelos, sistemas y algoritmos en el campo de la programación. Este método inspecciona periódicamente el producto que se está construyendo, estableciendo prioridades y especificaciones de entrega con valor para el usuario final. La figura 1 muestra el proceso incluido en la metodología de desarrollo ágil de Scrum.

Figura 1. Proceso de la metodología Scrum



Fuente: Elaboración propia. Tomada de Urteaga (2018) Aplicación de la Metodología Scrum

El proceso de metodología con Scrum permitió el desarrollo del trabajo mediante Sprints con fases semanales durante los meses de noviembre y abril, realizando las debidas inspecciones del modelo a construir y ajustes a la red social con el fin de evidenciar la correcta aplicación del algoritmo de similitud del coseno. Según, Ticona (2015) los flujos de trabajo o Sprints se denominan ciclos iterativos, en donde se realizan mejoras al producto que se integra a Scrum, en el Sprint el producto se codifica y demuestra, con oportunidad de evolucionar en el desarrollo de arquitectura y diseño (p.11).

Tabla 1. Sprints incluidos en el desarrollo de la metodología Scrum

mes Sprints	noviembre (2020)	diciembre (2020)	enero (2021)	febrero (2021)	marzo (2021)	abril (2021)
Creación de la red social, bases de datos, requisitos específicos						
Arquitectura de software, lenguaje de programación, framework, fórmula matemática del algoritmo, sistema de recomendación						
Mantenimiento e inspección						
Implementación del aplicativo						
Evidencia y efectividad del algoritmo						
Revisión y ajustes						

2.3 Fases de la investigación

- Características y funcionalidad de la red social: descripción de la red social, capturas de acceso y búsqueda de información por parte del usuario.
- Arquitectura de software para la aplicación del algoritmo similitud de coseno: lenguaje de programación, framework base de datos, fórmula matemática y sistema de recomendación.
- Modelo matemático de similitud de coseno para reconocer gustos de usuario en una red social: evidencia de la efectividad del modelo, capturas y análisis de la información.

2.4 Recursos

Los recursos empleados para el desarrollo la aplicación del modelo matemático similitud de coseno para identificar gustos de usuario en una red social se establecen en la tabla 2.

Tabla 2. Recursos para para el desarrollo del aplicativo

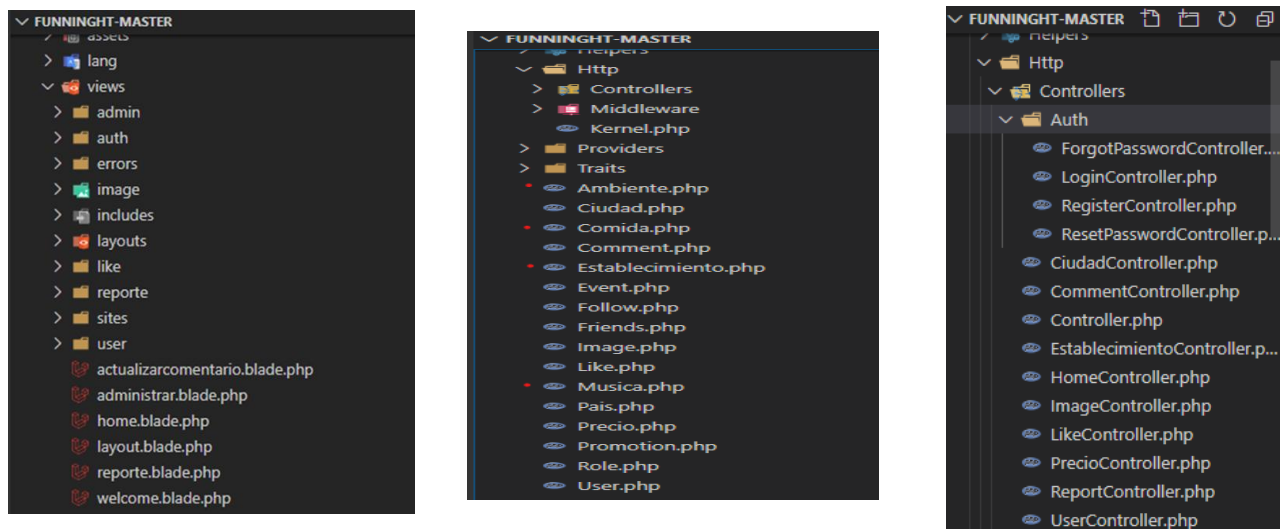
Red social	API del algoritmo
Framework: Laravel	Framework: Flask
Lenguaje: php	Lenguaje: Phyton
Base de datos: MySQL	Base de datos: MySQL

3. RESULTADOS Y DISCUSIÓN

3.1 Características y funcionalidad de la red social

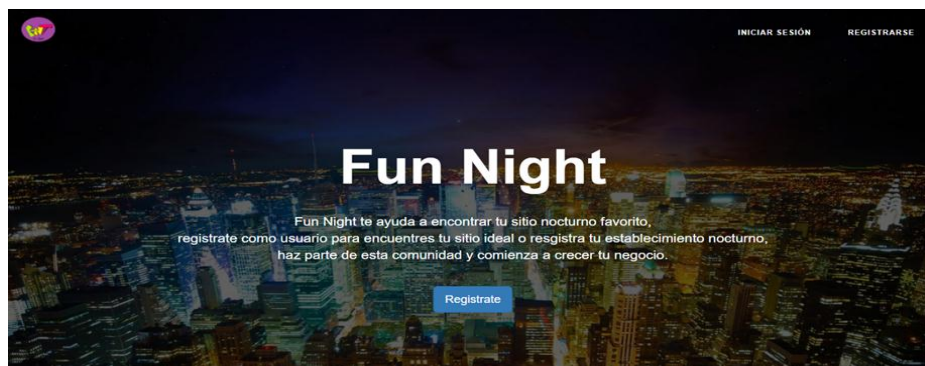
Las redes sociales en el ámbito tecnológico se conceptúan como sitios web integrales que permiten a los usuarios crear perfiles, carga de recursos personales y conectarse con múltiples redes a través de internet, actualmente representan uno de los medios más utilizados para acceder a la compra de servicios, productos e intercambio de información de diversa índole, brindando a los usuarios una rápida comunicación electrónica de contenido (Umit y Alatas, 2019). En ese sentido, la presente red social funciona como un sitio web que brinda al usuario la posibilidad de conocer la oferta gastronómica, cultural y de entretenimiento nocturno en la ciudad de Santiago de Cali, con el propósito de agilizar el proceso de búsqueda de tendencias de consumo de servicios. Es importante destacar, que en la red social se pone a prueba la aplicación del algoritmo de similitud de coseno con usuarios reales pertenecientes a la base de datos del sitio. Teniendo en cuenta lo anterior, se presentan la funcionalidad y características de la red social. La figura 2 describe el modelo vista del controlador de la red social con las funciones de la red social, desarrollado con Laravel y Php.

Figura 2. Modelo de vista del controlador de la red social



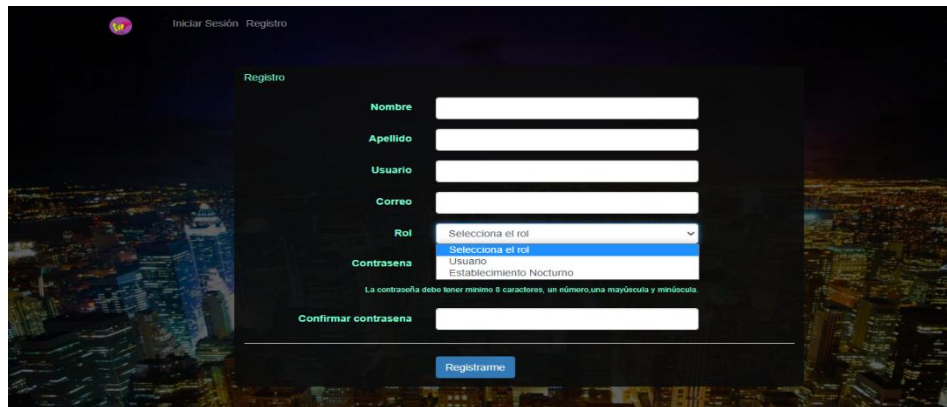
La figura 3 muestra la página de inicio o registro, destacando de fondo una imagen de ciudad, el nombre de la red social “Fun Night”, mensaje de entrada y finalidad del sitio web.

Figura 3. Entrada principal para inicio de sesión o registro principal



La figura 4 destaca los datos que debe ingresar el usuario para el registro en la red social como nombre, apellido, nickname de usuario, correo electrónico, rol, contraseña y confirmación de contraseña.

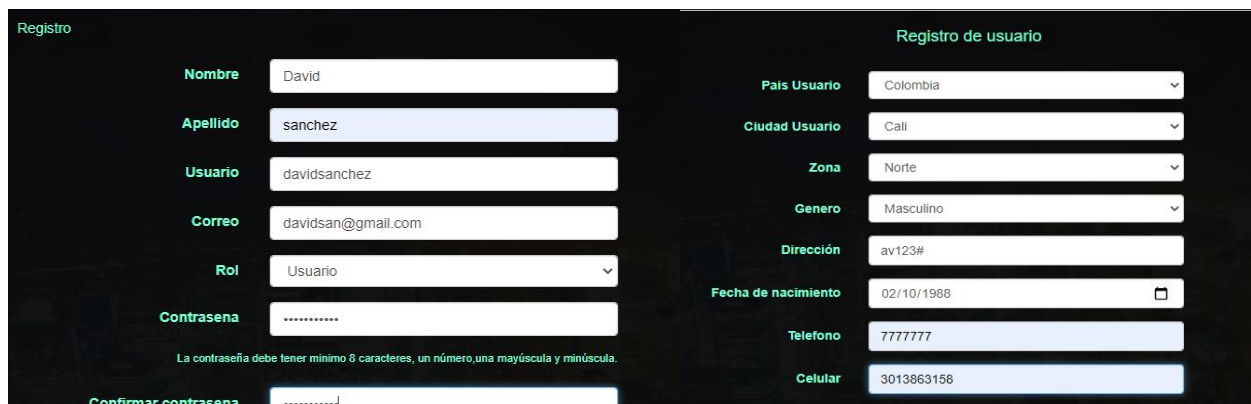
Figura 4. Registro sin datos



The screenshot shows a registration form titled 'Registro' on a dark background with a city skyline. The form includes input fields for 'Nombre', 'Apellido', 'Usuario', 'Correo', 'Rol' (with a dropdown menu), 'Contraseña', and 'Confirmar contraseña'. A blue 'Regístrate' button is at the bottom. A small note below the password field states: 'La contraseña debe tener mínimo 8 caracteres, un número, una mayúscula y minúscula.'

La figura 5 presenta la página de registro con datos reales de usuario, en este proceso se adiciona el país, ciudad, zona, genero, dirección, fecha de nacimiento y número telefónico.

Figura 5. Registro con datos



The screenshot shows the same registration form, but now with data entered. The left side is titled 'Registro' and the right side is titled 'Registro de usuario'. The data entered is as follows:

Registro	Registro de usuario
Nombre: David	País Usuario: Colombia
Apellido: sanchez	Ciudad Usuario: Cali
Usuario: davidsanchez	Zona: Norte
Correo: davidsan@gmail.com	Genero: Masculino
Rol: Usuario	Dirección: av123#
Contraseña: [oculto]	Fecha de nacimiento: 02/10/1988
Confirmar contraseña: [oculto]	Telefono: 7777777
	Celular: 3013863158

A small note below the password field states: 'La contraseña debe tener mínimo 8 caracteres, un número, una mayúscula y minúscula.'

En la figura 6, se reconocen los datos de los cuatro gustos principales del usuario que serán tomados en cuenta para realizar la funcionalidad principal del algoritmo de similitud de coseno como tipo de establecimiento, comida, ambiente y música preferida. Este proceso destaca que frente a los gustos del usuario, la red social posibilita la escogencia de diferentes alternativas, siendo más específica la elección del sitio y recomendación.

Figura 6. Gustos reconocidos por parte del usuario en la red social

The screenshot shows a user preference form with the title "Háblanos de ti. Cuáles son tus gustos?". It contains four main sections, each with a dropdown menu:

- Tipo de establecimiento preferido:** The dropdown menu is open, showing options: Bar, Restaurante, Taberna, Discoteca, restaurante - Bar, and Discoteca - Bar.
- Comida preferida:** The dropdown menu is open, showing options: Comida rápida, Comida americana, Comida Gourmet, Comida Mexicana, Comida Colombiana, Licor, Licor y Cócteles, Varios (Comida rápida, Comida Gourmet), Cócteles, Jugos y Betidos, Varios (Comida, Licor), and Cerveza.
- Ambiente preferido en un sitio nocturno:** This section is visible but the dropdown menu is not open.
- Musica que te gusta escuchar:** The dropdown menu is open, showing options: Bar, Restaurante, Taberna, Discoteca, restaurante - Bar, and Discoteca - Bar.

La figura 7 demuestra los cuatros gustos seleccionados por el usuario real frente al tipo de establecimiento, comida, ambiente y música preferida.

Figura 7. Gustos seleccionados por el usuario

The screenshot shows the same user preference form as in Figure 6, but with the selected options displayed in the dropdown menus:

- Tipo de establecimiento preferido:** Bar
- Comida preferida:** Comida rápida
- Ambiente preferido en un sitio nocturno:** Ambiente Clasico
- Musica que te gusta escuchar:** Rock

En la figura 8 se presentan las recomendaciones de la red social en función de los gustos establecidos por el usuario, a manera de lista la red social clasifica en primera instancia los sitios que cumplen con el 100% de las preferencias establecidas en la búsqueda, disminuyendo porcentualmente la compatibilidad hasta un 25% de similitud.

Figura 8. Lista de recomendación de los sitios al usuario

The screenshot shows a recommendation list titled "Sitios recomendados" with the subtitle "Según tus preferencias de tipo de establecimiento, comida, música y ambiente". It displays five user profiles with their names and similarity percentages:

- funn**: fun2 funnnn, Similaridad con gustos: 25 %
- funnighhhh**: funny night, Similaridad con gustos: 100 %
- route66**: ruta55 ruta66, Similaridad con gustos: 25 %
- a**: a a, Similaridad con gustos: 75 %
- rino**: javerino javier, Similaridad con gustos: 25 %

Each profile includes a "Ver Perfil" button.

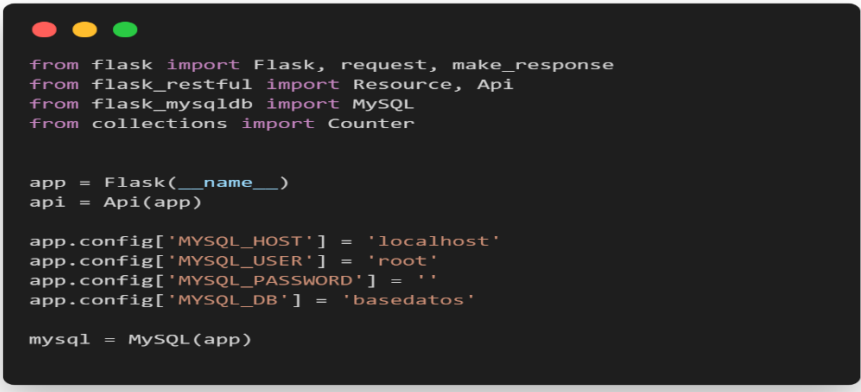
3.2 Arquitectura de software para la aplicación del algoritmo similitud de coseno en la red social

3.2.1 Lenguaje de programación

El lenguaje para la interfaz de programación del algoritmo de similitud de coseno utilizó Python, según Díaz y Becerra (2014) Python es lenguaje de programación multiparadigma, que tiene facilidades para la programación imperativa, funcional y orientada objetos, emplea una estructura dinámica con conteo de referencias para la gestión de la memoria (p.18). Como características que destacan a Python se reconocen:

- Inclusión de un sistema de documentación.
- Presenta un intérprete o consola para someter a prueba capacidades específicas del lenguaje, sin la necesidad de crear un módulo para este.
- Módulos que cubren necesidades esenciales de programación como estructura de la información, funciones numéricas, servicio para el sistema operativo y gestión de datos.
- Implementa la librería estándar en el lenguaje C, lo que permite que sus funciones básicas sean eficientes y productivas.

Figura 9. Código principal del algoritmo hecho en Python



```
from flask import Flask, request, make_response
from flask_restful import Resource, Api
from flask_mysql import MySQL
from collections import Counter

app = Flask(__name__)
api = Api(app)

app.config['MYSQL_HOST'] = 'localhost'
app.config['MYSQL_USER'] = 'root'
app.config['MYSQL_PASSWORD'] = ''
app.config['MYSQL_DB'] = 'basedatos'

mysql = MySQL(app)
```

3.2.2 Framework

El framework de la interfaz de programación del algoritmo de similitud de coseno empleó Flask como entorno de trabajo que permite la creación de aplicaciones web de manera ágil, con un reducido número de líneas de código, fundamentada con especificaciones de la Web Server Gateway de Werkzeug y templates Jinja2, a su vez cuenta con la licencia Berkeley Software Distribution. Frente a la estructura de Flask, Saavedra (2018) menciona que los valores predeterminados se componen de static y templates, debiendo estar ambos en el mismo directorio donde existe la aplicación, como beneficios de Flask permite registrar nuevos usuarios, recolección rápida de nuevas ideas, pudiendo estar completamente en un solo archivo o fraccionado en varios archivos dentro de un paquete.

3.2.3 Formula matemática

La operación del algoritmo de similitud del coseno mide el ángulo entre dos vectores de cada par de usuarios o ítems, relacionando similitud en aquellos vectores que tienen la misma orientación, se usa en espacio positivo y el resultado se establece en $[0,1]$. Proporcionando un valor de 1 si el ángulo comprendido es cero, esto es si ambos vectores apuntan a un mismo lugar, cualquier ángulo que se determine entre los vectores el coseno arrojará valores inferiores a 1. Hoi y Tieu (2018) destacan que en las métricas de calidad, las medidas del algoritmo de similitud del coseno son constantemente utilizadas, ya que aportan atributos al rendimiento de servicios web (p.33).

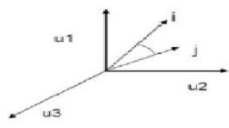
Figura 10. Fórmula matemática del algoritmo similitud del coseno

$A = [\text{ROCK}, \text{PIZZA}, \text{CLASICO}, \text{CERVEZA}]$
 $B = [\text{ROCK}, \text{CARNE}, \text{JUVENIL}, \text{CERVEZA}]$
 $U = [\text{ROCK}, \text{PIZZA}, \text{CLASICO}, \text{CERVEZA}, \text{CARNE}, \text{JUVENIL}]$

$A = [1, 1, 1, 1, 0, 0]$
 $B = [1, 0, 0, 1, 1, 1]$

$$\text{similarity} = \cos(\theta) = \frac{A \cdot B}{\|A\| \|B\|}$$

$$\text{COS}(\theta) = \frac{1+0+0+1+0+0}{2.2 \times 4} = \frac{2}{8} = \frac{1}{4} = 0.25$$



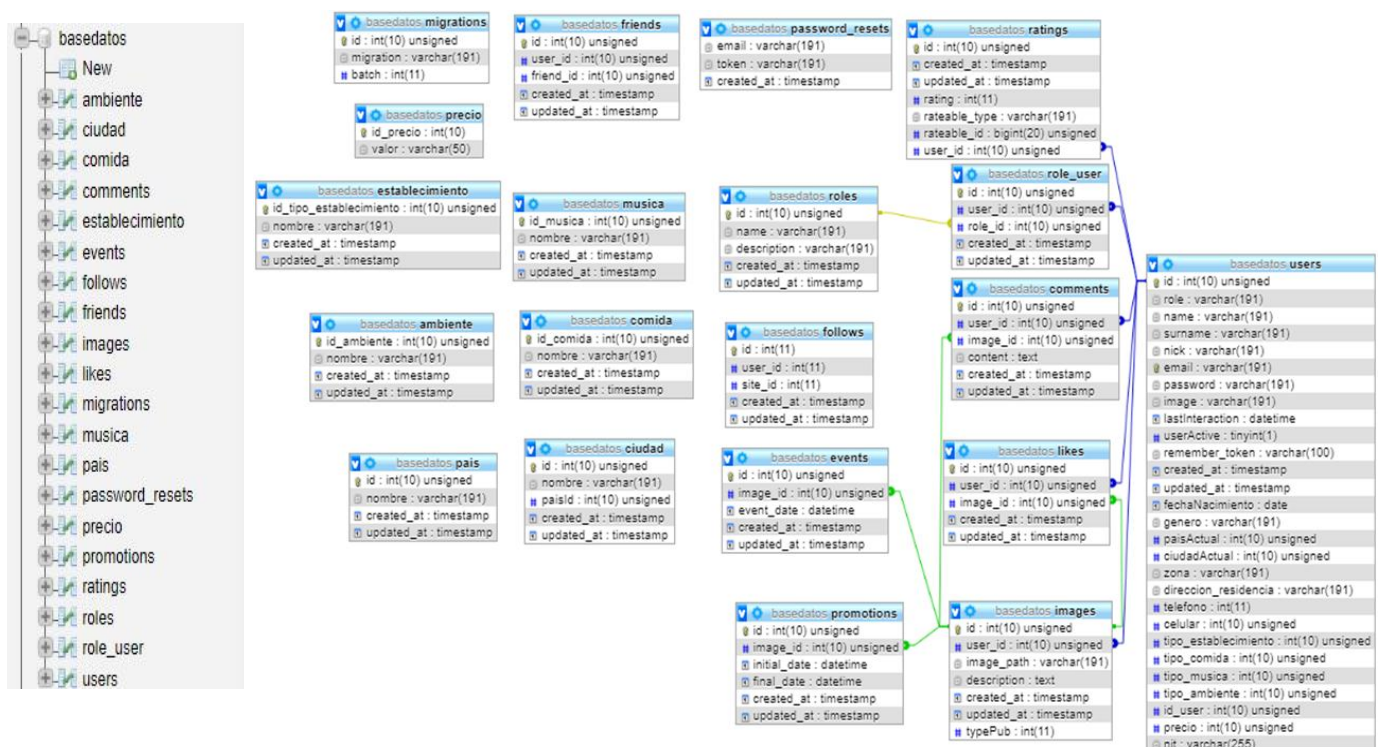
El algoritmo de similitud de coseno ejecuta el cálculo matemático para determinar niveles de semejanzas entre un conjunto de datos, por ende, en la programación se seleccionan los gustos o preferencias para formar posibles afinidades entre usuario e ítems, de esta manera se establece una recomendación del restaurante o sitio nocturno de interés para el usuario de la red social.

3.2.4 Base de datos

La base de datos es diseñada y administrada mediante el procesador de bases de datos MySQL, que suministra la capacidad de cifrar la información con el algoritmo, según Flores (2018) MySQL es un sistema de gestión de base de datos que brinda los componentes para agregar, acceder y gestionar la información almacenada en una base de datos (p.66). Dentro de las características se reconocen los siguientes aspectos:

- Organización de campos de texto.
- Maneja diferentes herramientas para el proceso de portabilidad.
- Permite suficiente almacenamiento de datos.
- Presenta el tipo de Licencia GPL para software libre.

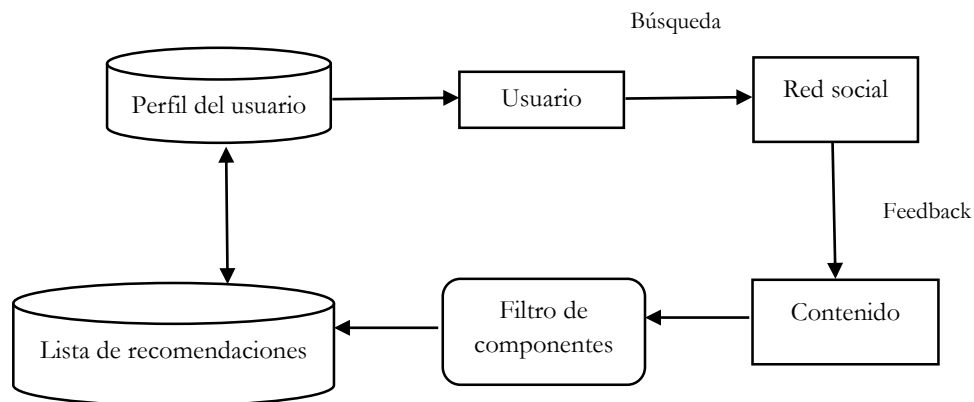
La figura 11 establece la base de datos con MySQL utilizada en el aplicativo.

Figura 11. Base de datos

3.2.5 Sistema de recomendación filtrado basado en contenido

El sistema de recomendación utilizado se denomina filtrado basado en contenido (Content-based Filtering), utiliza características y atributos de los elementos de búsqueda para sugerir nuevos productos o servicios que le pueden interesar al usuario en la red social (Marinas, 2014, pág. 7). De acuerdo con Son y Beoum King (2017) el sistema debe recomendar elementos relevantes para cada usuario, siendo específica la sugerencia, puesto que el filtrado basado en contenido no utiliza información previa sobre otros usuarios como comentarios, calificaciones o ratings sobre el ítem. Por ende, el sistema de recomendación basado en contenido selecciona los elementos en función de la correlación entre el contenido suministrado por la red social y los gustos del usuario, en ese sentido, el sistema de recomendación bajo el modelo de filtrado basado presenta el siguiente esquema:

Figura 12. Esquema del sistema de recomendación de filtrado basado en contenido aplicado en la red social



Nota. Elaboración propia con elementos del modelo de Someren (2013)

La figura 12 explica que el perfil se relaciona con los datos suministrados por el usuario en la búsqueda de la información sobre restaurantes y sitios nocturnos en la red social, estableciendo la lista recomendaciones de acuerdo con el filtrado basado en contenido, de los lugares con un grado de precisión exacto basado en las preferencias y gustos del cliente.

3.3 Funcionamiento del modelo matemático de similitud de coseno para reconocer gustos de usuario en una red social

Para identificar el funcionamiento del modelo matemático de similitud de coseno en el reconocimiento de gustos de usuario en una red social, se presentan las capturas en formato JavaScript Object Notation (JSON) que evidencian el proceso, este tipo de formato se utiliza para el análisis y procesamiento confiable de datos entre usuarios y servidor. En ese sentido, la figura 13 evidencia los sitios recomendados por el algoritmo de acuerdo a los índices de similitud, teniendo en cuenta el tipo de establecimiento, comida, ambiente y música preferida, por tanto, la lista arroja los lugares de entretenimiento nocturno y del sector gastronómico recomendado en la ciudad de Santiago de Cali, con similitud del 100% y mínimo del 25% de acuerdo con las preferencias del cliente, en este aspecto Rivas y Corona (2015) destacan que ejecutando el algoritmo de similitud de coseno se logra un cálculo con precisión de la información que se busca, pudiendo utilizarse para la medición de la coherencia entre los datos.

Figura 13. Sitios recomendados al usuario aplicando el algoritmo de similitud de coseno

```

- {
  sitio_id: 10,
  nick: "funnightrt",
  name: "funny",
  surname: "night",
  image: "1548447032fun_perfil.jpg",
  similaridad: 100
},
- {
  sitio_id: 19,
  nick: "a",
  name: "a",
  surname: "a",
  image: "",
  similaridad: 75
},
- {
  sitio_id: 13,
  nick: "funbarrock",
  name: "funbar",
  surname: "rock",
  image: "",
  similaridad: 50
},
- {
  sitio_id: 15,
  nick: "funn",
  name: "fun2",
  surname: "funnnn",
  image: "",
  similaridad: 25
},
- {
  sitio_id: 23,
  nick: "route66",
  name: "ruta55",
  surname: "ruta66",
  image: "",
  similaridad: 25
},
- {
  sitio_id: 18,
  nick: "rino",
  name: "javerino",
  surname: "javier",
  image: "",
  similaridad: 25
},
- {
  sitio_id: 21,
  nick: "c",
  name: "c",
  surname: "c",
  image: "",
  similaridad: 25
}

```

La lista de recomendaciones al usuario reconoce en la red social, al sitio 2night con un índice de similitud del 100%, Funbarrock 50%, Route 66 y Javerino con el 25%. Las siguientes capturas demuestran la extracción de datos que se usaron para el funcionamiento del algoritmo de similitud de coseno, función de similitud de coseno- vectorización de los datos y clase que contiene los modelos de los gustos evaluados y ruta.

Figura 14. Extracción de datos para el funcionamiento del algoritmo de similitud de coseno

```

def recomendar(caracteristicas):
    cur = mysql.connection.cursor()
    cur.execute(
        """SELECT users.id AS id_sitio, users.nick AS nick, users.name AS
        nombre, users.surname AS surname, users.image AS image,
        establecimiento.nombre AS tipo_establecimiento,
        comida.nombre AS tipo_comida,
        musica.nombre AS tipo_musica, ambiente.nombre AS
        tipo_ambiente
        FROM users
        INNER JOIN establecimiento
        ON establecimiento.id_tipo_establecimiento = use
        rs.tipo_establecimiento
        INNER JOIN comida
        ON comida.id_comida = users.tipo_comida
        INNER JOIN musica
        ON musica.id_musica= users.tipo_musica
        INNER JOIN ambiente
        ON ambiente.id_ambiente = users.tipo_ambiente
        WHERE role=3"""
    )

    similaridades = []
    for row in cur.fetchall():
        cosine = cosine_similarity(caracteristicas, row[5:])
        if cosine != 0:
            similaridades.append({'sitio_id': row[0], 'nick':
            row[1], 'name': row[2], 'surname': row[3], 'image': row[4],
            "similaridad": cosine*100})

    return sorted(similaridades, key=lambda k: k['similaridad'],
    reverse=True)

```

Figura 15. Función de similitud de coseno y vectorización de los datos

```
def cosine_similarity(c1, c2):  
    # count word occurrences  
    c1_vals = Counter(c1)  
    c2_vals = Counter(c2)  
  
    # convert to word-vectors  
    words = list(c1_vals.keys() | c2_vals.keys())  
    c1_vect = [c1_vals.get(word, 0) for word in words]  
    c2_vect = [c2_vals.get(word, 0) for word in words]  
  
    # find cosine  
    len_c1 = sum(av*av for av in c1_vect) ** 0.5  
    len_c2 = sum(bv*bv for bv in c2_vect) ** 0.5  
    dot = sum(av*bv for av, bv in zip(c1_vect, c2_vect))  
    cosine = dot / (len_c1 * len_c2)  
    return cosine
```

Figura 16. Clase que contiene los modelos de los gustos evaluados y ruta

```
class Recomendacion(Resource):  
    def get(self):  
        required_args = ['establecimiento', 'comida', 'musica',  
                        'ambiente']  
        for required_arg in required_args:  
            if not required_arg in request.args:  
                return make_response({'error':  
                    "falta el argumento " + required_arg}, 400)  
  
        caracteristicas = [value for value in  
            request.args.values()]  
        return recomendar(caracteristicas)  
  
api.add_resource(Recomendacion, '/recomendacion/')  
  
if __name__ == '__main__':  
    app.run(debug=True)
```

4. CONCLUSIONES

El funcionamiento del modelo matemático de similitud de coseno para reconocer gustos de usuario en una red social, al ejecutar el algoritmo logró un conteo exacto de los datos del 100% al 25%, teniendo en cuenta la afinidad de las preferencias establecidas por el cliente en la búsqueda; la integración del algoritmo de coseno al sistema web con el modelo de red social permitió recomendar sitios de interés teniendo en cuenta la información suministrada, dando un uso de tipo aplicado a tendencias actuales como son las redes sociales y los algoritmos basado en sistemas de recomendación que en el presente caso utilizo el filtrado basado en contenido, reconociendo características y atributos de los elementos de búsqueda para sugerir productos o servicios que le pueden interesar al cliente. En consecuencia, se ha cumplido con la función principal del aplicativo, la cual consiste en sugerir un establecimiento comercial sobre los gustos personales que cualquier usuario puede tener dentro de un escenario comercial con tendencias de consumo. Desde esa perspectiva, es importante mencionar que la presente investigación abre la posibilidad de generar proyectos futuros interés que desde el interés académico y práctico puedan surgir, mediante la implementación una amplia gama de algoritmos que se basan en principios y técnicas para la generación de similitudes, adaptables a diferentes escenarios como plataformas electrónicas de venta, redes sociales, sitios de streaming digital y entretenimiento audiovisual.

La red social se ha establecido como un sitio web, que ofrece al usuario alternativas de búsqueda del sector gastronómico y de entretenimiento nocturno en la ciudad de Santiago de Cali, facilitando de forma ágil el acceso a la información sobre los tipos de establecimientos, opciones de restaurante y ambientes musicales, teniendo en cuenta que el registro y consulta en línea brinda a los usuarios una rápida comunicación electrónica de contenido, por ende, se ha desarrollado una red social que posibilita la escogencia de diferentes alternativas de consumo, siendo específica la recomendación del sitio para mayor satisfacción y mejor toma de decisiones por parte del cliente.

Frente a la arquitectura del software para la aplicación del algoritmo similitud de coseno en la red social, constituida en el lenguaje de programación, framework, base de datos por medio de Python, Flask y MySQL ha propuesto una estructura de funcionalidad tecnológica con una serie de componentes hacia un fin específico, por cuanto el desarrollo del aplicativo promueve un producto de utilidad al servicio de una necesidad o problemática, proceso que se vincula a la línea de investigación de desarrollo de sistemas informáticos que emplea como recurso principal el manejo, almacenamiento, procesamiento y transformación de datos, para obtener un producto final que suministra a los diferentes usuarios retroalimentación fundamental, la cual se valora si la información obtenida se adecua a un resultado que genere beneficios y eficiencia, en ese sentido, la programación ha seleccionado los gustos o preferencias para formar posibles afinidades entre usuario y contenido, de esta manera se ha generado una recomendación del restaurante o sitio nocturno de interés.

REFERENCIAS

- Bayona, J., & Pardo, O. (2015). El papel de la Ingeniería de Software en el desarrollo de aplicaciones. *Tecnología, Investigación y Academia*, 8(33), 1-23.
- Belazzoug, M., & Brahimi, M. (2020). An improved sine cosine algorithm to select features for text categorization. *Journal of King Saud University - Computer and Information Sciences*, 30(21), 454-464.
- Belsky, G. (2019). Why Text Mining May Be The Next Big. *Technology & Media*, 6(19), 20-33.
- Díaz, Y., & Becerra, R. (2014). El lenguaje de programación Python. *Revista Ciencias*, 4(59), 10-25.
- Flores, E. (2018). Implementación de una base de datos heterogénea distribuida entre los SGBD ORACLE, MySQL y PostgreSQL con replicación. *INNOVA Research Journal*, 3(16), 59-76.
- Ghobar, E. (2019). Un sistema de recomendación basado en perfiles generados por agrupamiento y asociaciones. *International Journal of Scientific & Engineering Research*, 22(5), 45-68.
- Global Nielsen Company. (2018). Estudio sobre tendencias de comida fuera del hogar. *Revista economía y mercadeo*, 12(7), 1-22.
- Gonzalez, M. (2016). Análisis e implementación de un sistema de recomendación para la lista de la compra. *Revista computo y tecnología*, 41(4), 18-36.
- Harbir, R. (2020). Movie Recommendation System using Cosine Similarity. *International Journal of Engineering and Advanced Technology (IJEAT)*, 9(76), 12-31.
- Hoda, R., & Salleh, N. (2018). The Rise and Evolution of Agile Software Development. *IEEE Software Journal*, 23(30), 102-123.
- Hoi, C., & Tieu, C. (2018). Autonomic Ranking and Selection Of Web Services by Using Single Value decomposition Technique. *QoS-Based*, 7(20), 30-49.
- Ibrahim, H., Husnu, S., & Kuo-Chi, F. (2020). Predictive self-learning content recommendation system for multimedia contents. *System and Tech*, 32(9), 110-127.
- Jabba, D., & Mejía, A. (2017). Influencia de la Ingeniería de Software en los Procesos de Automatización Industrial. *Revista información tecnológica*, 11(8), 45-67.
- Laudon, K. (2018). *Management information system*. India: Pearson Education.
- Lozada, J. (2018). Investigación Aplicada: Definición, propiedad intelectual e industria. *Revista Tics y sociedad*, 6(3), 22-45.
- Lozano, C., & Fuentes, F. (2017). Una metodología de recomendación al cliente bancario a partir del historial de reclamaciones bancarias. *Revista Atlántica de ciencia y Economía*, 1(23), 1-19.
- Marinas, S. (2014). *Suite de algoritmos de recomendación en aplicaciones reales*. Barcelona: Estudios de ingeniería y computación aplicada.
- Mendoza, G., & Pérez, M. (2019). Métricas de similaridad y evaluación para sistemas de recomendación de filtrado colaborativo. *RITI Journal*, 5(44), 1-22.
- Rivas, C., & Corona, V. (2015). Metodologías actuales de desarrollo de software. *Revista Innovación*, 2(5), 980-986.
- Saavedra, E. (2018). Flask en el ámbito de los framework. *Libre atix*, 6(32), 45-67.
- Saha, S., & Ghosh, M. (2020). Feature Selection for Facial Emotion Recognition Using Cosine Similarity-Based Harmony Search Algorithm. *Applied Sciences*, 10(5), 28-42.
- Someren, M. (2013). Using Content-Based Filtering for Recommendation. *NetlinQ journal*, 9(65), 33-50.

- Son, J., & Beoum King, S. (2017). Content-based filtering for recommendation systems using multiattribute networks. *Expert Systems With Applications Journal*, 89(34), 404-416.
- Tang, E. (2019). A quantum-inspired classical algorithm for recommendation systems. *Journal of theory of computing*, 3(56), 217-240.
- Texier, J., & De Guisti, M. (2019). El desarrollo de software dirigido por modelos en los repositorios institucionales. *Revista internacional Dyna*, 6(8), 186-202.
- Ticona, F. (2015). Metodología scrum para el desarrollo de software y gestión de proyectos en las pequeñas y medianas empresas. *Revista Científica de Investigación Andina*, 45(21), 8-20.
- Umit, C., & Alatas, B. (2019). A new direction in social network analysis: Online social network analysis problems and applications. *Elsevier*, 535(34), 212-234.
- Urteaga, A. (2018). Aplicación de la Metodología Scrum. *Revista de ingeniería de sistemas de la Universidad de Madrid*, 5(12), 12-30.