

**PROTOTIPO DE SISTEMA DE RECOMENDACIÓN DE INFORMACIÓN
SOBRE LAS MICROEMPRESAS UBICADAS EN LOS ALREDEDORES
DE LA FACULTAD TECNOLÓGICA DE LA UNIVERSIDAD DISTRITAL
SOPORTADO EN UN PORTAL WEB**

CAMILO ANTONIO RAMÍREZ MORALES

**UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ DE CALDAS
FACULTAD TECNOLÓGICA
TECNOLÓGICA EN SISTEMATIZACIÓN DE DATOS
BOGOTÁ D.C
2014**

**PROTOTIPO DE INFORMACIÓN SOBRE LAS MICROEMPRESAS
UBICADAS EN LOS ALREDEDORES DE LA FACULTAD
TECNOLÓGICA DE LA UNIVERSIDAD DISTRITAL SOPORTADO EN UN
PORTAL WEB.**

CAMILO ANTONIO RAMÍREZ MORALES

CÓDIGO: 20111078094

**Proyecto presentado como requisito para optar por el título de Tecnólogo en
Sistematización de datos**

**TUTORES
GLORIA ANDREA CAVANZO NISSO**

**GRUPOS DE INVESTIGACIÓN METIS
UNIVERSIDAD DISTRITAL FRANCISCO JOSÉ
DE CALDAS**

Nota de Aceptación:

Firma del Tutor

Firma del Jurado

Firma del Jurado

Bogotá D.C. 16 de Octubre de 2014

CONTENDIDO

RESUMEN.....	10
INTRODUCCIÓN.....	11
1 FASE DE DEFINICIÓN, PLANEACIÓN Y ORGANIZACIÓN	14
1.1 TÍTULO.....	14
1.2 TEMA.....	14
1.3 PLANTEAMIENTO DEL PROBLEMA	14
1.4 JUSTIFICACIÓN	15
1.5 SOLUCIÓN TECNOLÓGICA	16
1.6 OBJETIVOS.....	17
1.6.1 <i>OBJETIVO GENERAL</i>	17
1.6.2 <i>OBJETIVOS ESPECÍFICOS</i>	17
1.7 ALCANCES	17
1.8 DELIMITACIONES.....	18
1.8.1 <i>DELIMITACIÓN GEOGRÁFICA</i>	18
1.8.2 <i>DELIMITACIÓN TEMPORAL</i>	18
1.8.3 <i>DELIMITACIÓN TEMÁTICA</i>	18
1.8.4 <i>DELIMITACIÓN TÉCNICA</i>	19
1.9 MARCO DE REFERENCIA	20
1.9.1 <i>MARCO TEÓRICO</i>	20
1.9.1.1 SISTEMAS DE RECOMENDACIÓN [6].....	22
1.9.1.2 Descomposición en Valores Singulares (SVD) [7].....	35
1.9.1.3 SVD aplicado al filtrado colaborativo [7]	36
1.9.1.4 Portal web.....	44
1.9.2 <i>MARCO HISTÓRICO</i>	45
1.9.3 <i>MARCO CONCEPTUAL</i>	49
1.9.4 <i>MARCO METODOLÓGICO</i>	50
1.10 FACTIBILIDAD	52
1.10.1 <i>FACTIBILIDAD TÉCNICA</i>	52
1.10.2 <i>FACTIBILIDAD ECONÓMICA</i>	52
1.10.3 <i>FACTIBILIDAD LEGAL</i>	52
1.10.4 <i>FACTIBILIDAD OPERATIVA</i>	53
1.11 CRONOGRAMA.....	54
2 MODELO DEL NEGOCIO.....	56
2.1 MODELO DE PROCESOS.....	56
2.1.1 <i>Modelo de procesos Sistema de establecimientos</i>	56
2.1.2 <i>Modelo de procesos Sistema de usuarios</i>	57
2.1.3 <i>Modelo de procesos sistema de Recomendación</i>	59
2.2 MODELOS DE DOMINIO	59
2.2.1 <i>Modelo de Dominio sistema de establecimientos</i>	60
2.2.2 <i>Modelo de Dominio sistema de usuarios</i>	60
2.2.3 <i>Modelo de Dominio sistema Recomendador</i>	61
2.3 GLOSARIO DE TÉRMINOS.....	61

3 FASE DE REQUERIMIENTOS	63
3.1 REQUERIMIENTOS FUNCIONALES.....	63
3.2 REQUERIMIENTOS NO FUNCIONALES.....	63
3.3 DEFINICIÓN DE ACTORES DEL SISTEMA.....	64
3.4 LISTA PRELIMINAR DE CASOS DE USO	64
3.4.1 <i>Sistema de usuarios</i>	64
3.4.2 <i>Sistema de Establecimientos</i>	65
3.4.3 <i>Sistema recomendador</i>	65
3.5 MODELOS DE CASOS DE USO	65
3.5.1 <i>Modelos de Casos de Uso del sistema de establecimientos</i>	65
3.5.2 <i>Modelos de Casos de Uso del sistema de usuarios</i>	66
3.5.3 <i>Modelos de Casos de Uso del sistema Recomendador</i>	67
3.5.4 <i>Modelos de Casos de Uso integrado</i>	67
3.5.5 <i>Documentación Casos de Uso</i>	67
4 FASE DE ANÁLISIS.....	71
4.1 DIAGRAMAS DE SECUENCIA.....	71
4.1.1 <i>Sistema de usuarios</i>	71
4.1.2 <i>Sistemas de establecimientos</i>	72
4.1.3 <i>Sistema recomendador</i>	72
4.2 DIAGRAMA DE ACTIVIDAD	73
4.2.1 <i>Sistema de usuarios</i>	74
4.2.2 <i>Sistema de establecimientos</i>	75
4.2.3 <i>Sistema recomendador</i>	75
4.3 DIAGRAMAS DE COMUNICACIÓN	76
4.3.1 <i>Sistema de usuarios</i>	76
4.3.2 <i>Sistema de establecimientos</i>	77
4.3.3 <i>Sistema recomendador</i>	77
4.4 DIAGRAMA DE ESTADO	78
4.4.1 <i>Sistema de usuarios</i>	78
4.4.2 <i>Sistema de establecimientos</i>	78
4.4.3 <i>Sistema recomendador</i>	79
5 FASE DE DISEÑO	80
5.1 LISTA PRELIMINAR DE CLASES	80
5.1.1 <i>Sistema de Usuarios</i>	80
5.1.2 <i>Sistema de establecimientos</i>	80
5.1.3 <i>Sistema recomendador</i>	82
5.1.4 <i>Lista integrada de clases</i>	83
5.2 RESPONSABILIDAD DE LAS CLASES	84
5.3 MODELO DE ANÁLISIS	87
5.4 MODELO LÓGICO	88
5.5 MODELO FÍSICO	89
5.6 MODELO DE BASE DE DATOS.....	90
5.6.1 <i>Modelo relacional</i>	90
5.6.2 <i>Modelo de datos de Redis</i>	91
5.6.3 <i>Diccionario de datos</i>	91
5.7 MODELO DE INTERFAZ	95

6 FASE DE IMPLEMENTACIÓN	96
6.1 DIAGRAMA DE COMPONENTES	96
6.2 DIAGRAMA DE PAQUETES	96
6.3 DIAGRAMA DE DESPLIEGUE	97
7 FASE DE PRUEBAS	98
7.1 PRUEBAS DE INTEGRACIÓN.....	98
7.1.1 <i>Sistema de usuarios</i>	98
7.1.2 <i>Sistema de establecimientos</i>	99
7.1.3 <i>Sistema recomendador</i>	101
7.2 PRUEBAS DE SISTEMA.....	102
7.3 PRUEBAS DEL ALGORITMO DE FILTRADO COLABORATIVO SVD	104
7.3.1 <i>ERROR MEDIO ABSOLUTO (MAE)</i>	104
7.3.2 <i>Capacidad de recomendación Coverage [50]</i>	106
7.3.3 <i>Tiempos de ejecución</i>	106
CONCLUSIONES	108
8 BIBLIOGRAFÍA	109
ANEXOS	114

INDICE DE TABLAS

Tabla 1. Solución tecnológica	16
Tabla 2. Herramientas tecnológicas.....	20
Tabla 3. Problemas de Regularización	40
Tabla 4. Regularización	41
Tabla 5. Pseudocódigo de RSVD	44
Tabla 6. Aplicación de la metodología	51
Tabla 7. Factibilidad económica.....	52
Tabla 8. Factibilidad legal	53
Tabla 9. Glosario de términos	62
Tabla 10. Definición de actores del sistema	64
Tabla 11. Lista preliminar Casos de Usos Sistema de usuarios	64
Tabla 12. Lista preliminar Casos de Usos Sistema de establecimientos	65
Tabla 13. Lista preliminar Casos de Usos Sistema recomendador.....	65
Tabla 14. Documentación Caso de uso Registro convencional.....	69
Tabla 15. Documentación caso de uso Filtrado por localización	70
Tabla 16. Documentación del caso de uso Calificar establecimiento	70
Tabla 17. Lista preliminar de clases del Sistema de usuarios	80
Tabla 18. Lista preliminar de clases del Sistema de establecimientos	82
Tabla 19. Lista preliminar de clases del Sistema recomendador	82
Tabla 20. Lista integrada de clases	84
Tabla 21. Responsabilidad de la Clase Model.....	84
Tabla 22. Responsabilidad de la Clase Form	85
Tabla 23. Responsabilidad de la Clase View	85
Tabla 24. Responsabilidad de la Clase SVD	85
Tabla 25. Responsabilidad de la Clase Autocomplete	86
Tabla 26. Responsabilidad de la Clase Solicitud	86
Tabla 27. Responsabilidad de la Clase Notificaciones	86
Tabla 28. Diccionario de datos.....	95
Tabla 29 . Pruebas sistema de usuarios	99
Tabla 30. Fallos Sistema de usuarios	99
Tabla 31. Pruebas sistema de establecimientos.....	101
Tabla 32. Fallos sistema de establecimientos	101
Tabla 33. Pruebas sistema recomendador	102
Tabla 34. Fallos sistema recomendador	102
Tabla 35. Pruebas de sistema	103
Tabla 36. Fallos de sistema	104

INDICE DE ILUSTRACIONES

Ilustración 1. Resumen censo Ciudad Bolívar	12
Ilustración 2. Diagrama general de un Sistema de recomendación [6]	24
Ilustración 3. Diferentes formas de captura de preferencia [6].....	25
Ilustración 4. Algoritmos basados en usuarios. [6].....	31
Ilustración 5. Algoritmos basados en productos.....	32
Ilustración 6. Modelo de procesos Gestión de establecimientos	56
Ilustración 7. Modelo de procesos Calificar establecimiento.....	56
Ilustración 8. Modelo de procesos Solicitud administración	57
Ilustración 9. Modelo de procesos Gestión de imágenes.....	57
Ilustración 10. Modelo de procesos Registro	58
Ilustración 11. Modelo de procesos Restablecer contraseña.....	58
Ilustración 12. Modelo de procesos Gestión de usuarios.....	58
Ilustración 13. Modelo de procesos Gestión de notificaciones	59
Ilustración 14. Modelo de procesos Recomendar	59
Ilustración 15. Modelo de dominio Sistema de establecimientos	60
Ilustración 16. Modelo de dominio Sistema de usuarios	60
Ilustración 17. Modelo de dominio Sistema recomendador	61
Ilustración 18. Modelo de casos de uso Sistema de establecimientos	66
Ilustración 19. Modelo de casos de uso Sistema de usuarios	66
Ilustración 20. Modelo de casos de uso Sistema recomendador	67
Ilustración 21. Modelo de casos de uso Integrado.....	67
Ilustración 22. Diagrama secuencia Registro convencional.....	71
Ilustración 23. Diagrama secuencia Crear establecimiento	72
Ilustración 24. Diagrama de secuencia Calificar establecimiento	73
Ilustración 25. Diagrama de actividad Recuperar contraseña.....	74
Ilustración 26. Diagrama de actividad Filtrar búsqueda categoría	75
Ilustración 27. Diagrama de actividad Computar	75
Ilustración 28. Diagrama de comunicación Consultar notificaciones	76
Ilustración 29. Diagrama de comunicación Modificar establecimiento (Administrador)	77
Ilustración 30. Diagrama de comunicación Obtener recomendación	77
Ilustración 31. Diagrama de estado Notificación	78
Ilustración 32. Diagrama de estado Establecimiento	78
Ilustración 33. Diagrama de estado Solicitud	79
Ilustración 34. Diagrama de estado Recomendación.....	79
Ilustración 35. Diagrama de clases de Análisis Sistema recomendador.....	87
Ilustración 36. Diagrama de clases Lógico Sistema recomendador.....	88
Ilustración 37. Diagrama de clases Físico Sistema recomendador.....	89

Ilustración 38. Modelo relacional.....	90
Ilustración 39. Modelo de datos de Redis	91
Ilustración 40. Modelo de interfaz	95
Ilustración 41. Diagrama de componentes.....	96
Ilustración 42. Diagrama de paquetes	96
Ilustración 43. Diagrama de despliegue.....	97

RESUMEN

Internet como tecnología ha creado nuevas relaciones entre consumidores y microempresas, permitiendo el surgimiento de una nueva forma de comercio a través de la Web. A pesar de las facilidades para consultar información en la Web surgen nuevos problemas derivados del exceso y dispersión de información como también el desconocimiento del uso de las TIC por parte de los usuarios. Como solución a esto se han implementado los sistemas de recomendación que tiene como objetivo asistir y facilitar a los usuarios en la selección de información de acuerdo a los requerimientos obtenidas a partir la interacción con el sistema.

Por otro lado, en muchas de las microempresas Colombianas hay un desconocimiento y desconcierto de las herramientas que brinda Internet para el crecimiento comercial. En particular cuando las microempresas, que por definición no superan los diez trabajadores, no suelen contar con los recursos necesarios para tener un departamento de sistemas que se encargue de la construcción y mantenimiento ya sea páginas o portales Web sobre la empresa, este es el caso en el que se encuentra la mayoría de las microempresas de la localidad de Ciudad Bolívar en Bogotá Colombia.

Este proyecto genera una solución a las dos anteriores necesidades mediante la implementación de un prototipo de sistema de recomendación que usa el método de descomposición singular de valores (SVD por sus siglas en inglés) propuesto por Simon Funk aplicado al filtrado colaborativo, esto será soportado en un portal web.

INTRODUCCIÓN

Desde el nacimiento del comercio y el intercambio de bienes, las empresas han creado técnicas para darse a conocer, captar la atención de clientes, sobrevivir y competir en el mercado para poder ofrecer sus productos o servicios que satisfagan las necesidades de los diferentes tipos de usuarios. [1]

Actualmente muchas de las empresas que ofrecen sus productos o servicios lo hacen a través de Internet. “Internet sin duda es la estrategia más común para el despegue de las empresas en el ámbito online” [2]. De igual forma, los usuarios se han adaptado a los novedosos medios de comunicación que ofrece la tecnología, buscando por dichos medios encontrar la información acorde a sus intereses y necesidades [3]. Esta realidad ha hecho que se produzca una competencia cada vez más reñida entre los comerciantes y empresarios, buscando así los medios más efectivos para asegurar su éxito dentro del mercado creciente.

A pesar de que tanto los usuarios como las microempresas han sabido adaptarse, hay todavía grandes problemas. Problemas como la abundancia de información derivada de diferentes fuentes provocando contradicciones y ambigüedades, generando así un desconocimiento para los usuarios sobre ¿En dónde buscar?, ¿Cómo buscar?, etc. y en el mismo sentido en las microempresas, generando cuestionamientos tales como ¿Dónde anuncio?, ¿Cómo anuncio? ¿Qué anuncio? Así mismo, en el peor de los casos, a pesar de la diversidad de herramientas, vemos que los usuarios y anunciantes no han logrado familiarizarse con las nuevas tecnologías provocando un desaprovechamiento de las mismas. En consecuencia tanto las microempresas como los usuarios (consumidores) se ven afectados. A este fenómeno suele llamársele como “El problema de la sobrecarga de información” [4] [5]

Debido a esta necesidad surgieron los sistemas de recuperación de información generalmente conocidos como buscadores que tomaron un papel importante como herramienta indispensable para buscar información en internet, empresas como Google o Bing son grandes ejemplos de lo mencionado. [6]

Sin embargo los usuarios muchas veces no saben, no pueden comunicar sus necesidades o simplemente no tienen el tiempo de emplear grandes búsquedas. Es posible que no se conozca en realidad lo que se quiere o simplemente que se pierda nuestra necesidad en los ríos de información que crecen continuamente en internet, impidiendo encontrar esa información valiosa que se requiere. [6] [7]

De esta forma vemos como los sistemas de recuperación de información ya no son suficientes y es una necesidad encontrar herramientas que se encarguen de

buscar por nosotros, que nos conozcan y que nos recomiendan un conjunto limitado de opciones acordes a nuestros intereses. Este es esencialmente el objetivo de los sistemas de recomendación [7]

Los sistemas de recomendación son uno de los métodos más destacados y que más popularidad ha ganado en los últimos años, en especial en el comercio electrónico. Los sistemas de recomendación, son sistemas inteligentes, que tienen la capacidad de aprender y recopilar las preferencias, gustos o necesidades de los usuarios y a partir de este conocimiento brindar una información limitada pero relevante que será de interés [8] para este. En palabras más sencillas un sistema de recomendación será una especie de filtro creado por el mismo usuario que solo dejara pasar información relevante, este filtro se amolda continuamente [6]. Existen diversos tipos de sistemas de recomendación, en este proyecto nos enfocaremos en los basados en filtrado colaborativo con el uso de técnicas de reducción de dimensionalidad.

El proyecto no solo se enfoca en el uso de los sistemas de recomendación sino también de su aplicabilidad a los establecimientos comerciales en la localidad de Ciudad Bolívar en Bogotá Colombia donde el sector predominante es el comercial y se registraron un total de 8293 unidades productivas dedicadas a la compra y venta de artículos y bienes. La segunda actividad económica es servicios con 4610 unidades productivas registradas. Las tiendas representan el 11% de las unidades productivas, seguida de la comercialización de productos y en tercer lugar las misceláneas. En la siguiente tabla se muestra un resumen del censo aplicado en la localidad de ciudad Bolívar y que muestra dichas cifras.

Tienda	1.963	11,17
Comercio	758	4,31
Miscelánea	412	2,34
Panadería	330	1,88
Confecciones	303	1,72
Cabinas telefónicas	243	1,38
Agrícola	230	1,31

Fuente: Censo empresarial Ciudad Bolívar. CCB y Alcaldía de Ciudad Bolívar.

Cálculos: Dirección de Estudios e Investigaciones de la CCB.

Ilustración 1. Resumen censo Ciudad Bolívar

El 97% de las unidades son microempresas, el mayor porcentaje de las empresas son unipersonal (88%) seguido por las unidades productivas de carácter familiar con (6%). La mayoría de las empresas (36%) tienen más de cinco años de

funcionamiento y el 33% corresponde a las empresas que llevan entre uno y cinco años de haberse conformado y estar en funcionamiento.

Por otro lado se observa que el 30% corresponde a las empresas que llevan menos de un año de haberse conformado y estar en funcionamiento. El 43% de las unidades tienen una concentración de ingresos 100.000 a 500.000 pesos al mes. El 78% de las unidades productivas afirmaron que donde más necesidad de capacitación hay es en el área de sistemas. El 92% de las unidades productivas afirmaron no tener equipo de cómputo, Es decir tan sólo el 8% dijo tenerlo. De estas ocho unidades empresariales tan sólo el 4% tiene posibilidad de conectividad. [9]

Teniendo en cuenta las necesidades anteriormente mencionadas, este proyecto se enfocará en el campo de recomendaciones de establecimiento comerciales mediante el uso de métodos de descomposición de matrices (SVD por sus siglas en inglés) que se encuentran ubicados en la localidad de Ciudad Bolívar, Bogotá Colombia. Así, de esta manera poder diseñar y desarrollar un sistema de recomendación el cual residirá en un portal web que permita agregar estos establecimientos comerciales por parte de cualquier usuario y al mismo tiempo ser consultado y recomendado.

1 FASE DE DEFINICIÓN, PLANEACIÓN Y ORGANIZACIÓN

1.1 TÍTULO

Prototipo de sistema de recomendación de información sobre las microempresas ubicadas en los alrededores de la Facultad Tecnológica de la Universidad Distrital soportado en un portal Web

1.2 TEMA

El desarrollo del proyecto aborda temas de Sistemas de recomendación de igual manera técnicas de filtrado colaborativo, descomposición singular de valores aplicada a los sistemas de recomendación, metodologías derivadas entre otros.

1.3 PLANTEAMIENTO DEL PROBLEMA

La ciudad de Bogotá es la ciudad más poblada de Colombia, cuenta con una población estimada de 7'363.782 habitantes. Según estudios de la Cámara de Comercio de Bogotá el 10 de Julio de 2012 estaban registradas 205.998 organizaciones de personas naturales y jurídicas ubicadas en Bogotá. De las cuales 166.378 son microempresas que representan el 80.77% de la participación en el mercado [10]. Por otro lado, en Ciudad Bolívar se registró un total de 3.665 empresas. Ya para el 2012 en la misma localidad se encontraron 4196 empresas activas en el sector, Lo que indica claramente el crecimiento comercial en la localidad de Ciudad Bolívar. [11]

A causa del crecimiento comercial en dicha localidad se ha generado una competencia entre las empresas del sector con el fin de darse a conocer, ganar clientes, adquirir ganancias y expandirse, siendo estas labores no tan sencillas. Debido a esto, las microempresas deben tener la capacidad de adaptarse a un mercado cambiante, competitivo y a nuevas formas de comercio.

Por otro lado abordamos el problema a partir de los usuarios los cuales se enfrenta al problema recurrente que se produce al momento de realizar una búsqueda en Internet [2], como el exceso de información, que aunque usualmente se realiza con la ayuda de sistemas de recuperación de información como Google o Bing terminan siendo insuficientes para soluciones enfocadas a lo particular. Este problema suele producirse muchas veces porque los usuarios no saben, les es difícil comunicar sus necesidades, no tienen el tiempo de emplear complejas búsquedas o simplemente desconocen que es lo que realmente quieren. [4] [5]

En conclusión, podemos generalizar el problema desde dos enfoques separados que convergen en un mismo punto; desde el enfoque de un emprendedor que es propietario de un establecimiento y necesita ofertar ya sea sus productos o servicios a aquellas personas que en verdad lo requieren, pero que no tiene los suficientes conocimientos o recursos para lograrlos, y por supuesto el enfoque del consumidor o cliente que requiere encontrar solución a sus necesidades de una manera fácil y amigable mediante una asistencia que ayude en la selección de posibles opciones derivadas del exceso de información, eliminando así información irrelevante y dejando solo aquella información ajustada a sus requerimientos.

1.4 JUSTIFICACIÓN

El mundo de los negocios está hoy en constante movimiento. Y tecnologías como Internet es uno de los elementos que está ofreciendo miles de nuevas oportunidades de negocios para las grandes, medianas, pequeñas y Micro empresas, que encuentran en este medio una nueva forma de ofertar y dar a conocer sus productos o servicios [8].

En las organizaciones de hoy se ha convertido en vital importancia actualizarse y entrar en la Web para comunicarse y cumplir sus objetivos, pero para darse a conocer se debe brindar información que sea atractiva para un cliente y que supla sus necesidades; es ahí donde surge la necesidad de diseñar y usar herramientas que sean capaces de integrar la información con el fin de brindar y obtener lo que se busca, es decir se necesita hacer uso de sistemas que filtren la información según los requerimientos y criterios de búsqueda de un usuario.

En la localidad de ciudad Bolívar se encuentra un gran número de empresas conformadas por personas que se dedican a labores informales o están desempleadas. El sector predominante en la localidad es el comercial, principalmente el comercio de tenderos, vendedores ambulantes y estacionarios y la venta de alimentos en panaderías, cafeterías, fruterías y fábricas. En orden de importancia, le sigue el sector de servicios, donde sobresalen los de peluquería, taller de mecánica, plomería y limpieza. En tercer lugar está el sector manufacturero, donde actividades como la zapatería, la carpintería, las confecciones y la ornamentación son las principales. Además muchas familias forman microempresas que producen artículos de cuero y productos artesanales. [12]

En este sentido, se contribuirá en el desarrollo comercial por medio del uso de tecnologías que permiten facilitar y mejorar la forma en que se relacionan los usuarios con la información de las organizaciones y con las nuevas formas de buscar y encontrar información en la Web. Así, lograr contribuir en el crecimiento de emprendedores facilitando las formas de darse a conocer como organización dentro

del mercado y el mundo de las tecnologías como nuevas alternativas de competición.

1.5 SOLUCIÓN TECNOLÓGICA

Se busca poder facilitar la interacción entre los clientes y microempresas que se encuentran ubicadas en la localidad de Ciudad Bolívar, Bogotá Colombia mediante el uso de tecnologías web y de bases de datos. A continuación las describimos:

Tecnología	Uso
Django web framework	Se usara como el <i>framework</i> de soporte dinámico del portal web del lado del servidor (<i>Back-end</i>)
Html5, CSS3, JavaScript	Se usaran para el soporte del lado del cliente (<i>Front-end</i>); Html5 se usa para la representación semántica del contenido, CSS3 para la creación de estilos de visualización y JavaScript para el contenido dinámico ejecutado en el cliente.
Bootstrap 3	Se usa para la aplicación de estilos de HTML mediante el uso de clases ya predefinidas.
Jquery framework	Como <i>framework</i> que facilite el trabajo con JavaScript.
NGINX	Como servidor web proxy.
Redis	Como base de datos de acceso rápido para almacenar las recomendaciones.
Postgresql	Como bases de datos relacional que almacenará la información del portal web.
PostGis	Como base de datos espacial que implementara elementos de ubicación.
Sqlite	Como base de datos relacional de desarrollo.
SpatiaLite	Como base de datos espacial de desarrollo.
Google maps	Como herramienta para presentar las ubicaciones de los establecimientos.

Tabla 1. Solución tecnológica

Las anteriores tecnologías se implantaran en un servidor permitiendo que el cliente acceda al portal web a través de una dirección IP, traducida a un dominio. En la conexión se intercambiarán datos entre el servidor y el cliente mediante la interacción del usuario con el portal web permitiendo conocer los requerimientos del usuario.

El usuario interactuará a través de un navegador web, el cual está presente en equipo de escritorio, portátil y dispositivos móviles, la interacción está centralizada en los formularios de la aplicación, en especial los formularios de calificación.

Al momento que el usuario realiza una interacción directa con el portal (Calificar en un rango de puntaje) el servidor se encargará de procesar mediante el algoritmo de SVD (Singular Value Descomposition) aplicado al filtrado colaborativo generando así una predicción de calificación a otros ítems, se almacenaran y posteriormente

se recomendaran los N (en nuestro caso serán 10) mejores ítems no puntuados cuando el usuario lo requiera.

Se desarrollaran e implementarán tres módulos; de usuarios, de establecimientos y de recomendación de establecimientos. El primero de ellos, el módulo de usuarios, se encargará de todo lo relacionado con registro, autenticación y permisos de usuarios. El segundo, el módulo de establecimientos, se encarga de la gestión y almacenamiento de establecimientos, y el tercero el de recomendación, se encarga de la captura de preferencia y de aplicar el algoritmo de recomendación para generar recomendaciones a los usuarios.

1.6 OBJETIVOS

1.6.1 OBJETIVO GENERAL

Diseñar y desarrollar un prototipo de sistema de recomendación de información sobre las microempresas ubicadas en los alrededores de la Facultad Tecnológica de la Universidad Distrital el cual estará soportado en un portal Web.

1.6.2 OBJETIVOS ESPECÍFICOS

- Conocer los requerimientos y necesidades actuales de las microempresas mediante la elaboración y aplicación de cuestionarios.
- Analizar la información recopilada determinando así el alcance y las necesidades del portal web.
- Diseñar y desarrollar los módulos de persistencia, acceso y establecimientos.
- Desarrollar el módulo de recomendación de información de acuerdo a los criterios de búsqueda del usuario y a los productos y servicios que ofrecen las microempresas del sector.

1.7 ALCANCES

- Contribuir y facilitar la interacción de los usuarios con la tecnología web en aspectos comerciales de la vida cotidiana.
- Involucrar a las microempresas empresas con tecnologías Web, contribuyendo así en su crecimiento y reconocimiento ante los consumidores.
- Facilitar a las organizaciones el acceso al comercio electrónico sin la necesidad de que tengan un servicio informático contratado.
- Permitir realizar consultas sobre establecimientos comerciales que ofrecen productos o servicios de interés de los consumidores en los alrededores de la Facultad Tecnológica de la Universidad Distrital.

- Diseñar e implementar módulos de registro e ingreso de los usuarios del sistema, de establecimientos y recomendación de establecimientos.

1.8 DELIMITACIONES

1.8.1 DELIMITACIÓN GEOGRÁFICA

El prototipo de sistema de recomendación de información sobre microempresas será implementado dentro de las instalaciones del grupo de investigación METIS. El enfoque del proyecto será en los establecimientos comerciales ubicados en los alrededores de la Facultad Tecnológica de la Universidad Distrital Francisco José de Caldas, localidad de Ciudad Bolívar, Bogotá Colombia.

1.8.2 DELIMITACIÓN TEMPORAL

Para el completo y correcto diseño y desarrollo del presente proyecto se ha estimado una duración en el tiempo de nueve (9) meses.

1.8.3 DELIMITACIÓN TEMÁTICA

La temática que abarca el desarrollo del prototipo es:

1. Sistemas de recomendación.

- Concepto.
- Tipos.
- Captura de preferencias.
- Técnica de descomposición de valores singulares (SVD) en el filtrado colaborativo.
 - Ventas y desventajas
 - SGD (Descenso del gradiente estocástico)
- Aplicaciones.

2. Establecimientos y microempresas.

- Conceptos.

Para el desarrollo del sistema recomendador, se usará el algoritmo de SVD aplicado al filtrado colaborativo, con las mejoras propuestas por Simón Funk. Lo anterior se realizará mediante el uso de las librerías de Pyrecsys¹. No se tratarán problemas propios del filtrado colaborativo como la manipulación de preferencias (*spam*), y escases de puntuaciones y escalabilidad.

¹ <http://ocelma.net/software/python-recommender-systems/build/html/index.html>

Por el lado de los establecimientos comerciales, no nos enfocaremos en el proceso de comprobación y validación de la propiedad de un establecimiento de un usuario, por lo tanto será visto como un proceso externo.

1.8.4 DELIMITACIÓN TÉCNICA

Para el desarrollo del prototipo de sistema de recomendación se requiere de lenguajes de programación y herramientas que permitan el desarrollo web y la elaboración de sistemas de recomendación, a continuación las describimos:

Software		
Tecnología	Descripción	Disponibilidad
Ubuntu server 14.04	Ubuntu es un sistema operativo basado en Linux y que se distribuye como software libre. [13]	Gratuita
Python 2.7	Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en una sintaxis que favorezca un código legible. Se trata de un lenguaje de programación multiparadigma, ya que soporta orientación a objetos, programación imperativa y, en menor medida, programación funcional. Es un lenguaje interpretado, usa tipado dinámico y es multiplataforma. Es administrado por la Python Software Foundation. [14]	Gratuita
Git	Git es un software de control de versiones diseñado por Linus Torvalds, pensando en la eficiencia y la confiabilidad del mantenimiento de versiones de aplicaciones cuando estas tienen un gran número de archivos de código fuente. [15]	Gratuita
Django framework 1.6.6	Django es un framework de desarrollo web de código abierto, escrito en Python, que respeta el paradigma conocido como Model Template View. La meta fundamental de Django es facilitar la creación de sitios web complejos. Django pone énfasis en el re-uso, la conectividad y extensibilidad de componentes, el desarrollo rápido y el principio No te repitas (DRY, del inglés Don't Repeat Yourself). Python es usado en todas las partes del framework, incluso en configuraciones, archivos, y en los modelos de datos. [16]	Gratuita
Bootstrap 3	Es una colección de herramientas para la creación de sitios y aplicaciones web. Este contiene creaciones base de CSS, HTML y Fuentes de texto, para botones, formularios, barras de navegación, etc. En junio de 2014 fue el proyecto número 1 de GitHub. [17]	Gratuita
Jquery 1.9	jQuery es una biblioteca de JavaScript, que permite simplificar la manera de interactuar con los documentos HTML, manipular el árbol DOM, manejar eventos, desarrollar animaciones y agregar interacción con la técnica AJAX a páginas web. jQuery es la biblioteca de JavaScript más utilizada. [18]	Gratuita

Celery	Celery es un proyecto opensource que permite la ejecución asíncrona de las tareas definidas en una cola de trabajos. [19]	Gratuita
RabbitMQ	RabbitMQ es un software de negociación de mensajes de código abierto, y entra dentro de la categoría de middleware de mensajería. Implementa el estándar Advanced Message Queuing Protocol (AMQP). El servidor RabbitMQ está escrito en Erlang y utiliza el framework Open Telecom Platform (OTP) para construir sus capacidades de ejecución distribuida y conmutación ante errores. [20]	Gratuita
Nginx	Nginx ("engine x") es un servidor HTTP y proxy inverso de alto rendimiento, y un servidor proxy para IMAP/POP3/SMTP. Nginx fue desarrollado por Igor Sysoev para Rambler.ru. Aunque aún se encuentra en una etapa beta, Nginx es conocido por su estabilidad, gran conjunto de características, configuración simple, y bajo consumo de recursos. [21]	Gratuita
Redis	Redis es un motor de base de datos en memoria, basado en el almacenamiento en tablas de hashes (clave/valor) pero que opcionalmente puede ser usada como una base de datos durable o persistente. Está escrito en ANSI C por Salvatore Sanfilippo. [22]	Gratuita
Postgresql	PostgreSQL es un SGBD relacional orientado a objetos y libre. [23]	Gratuita
PostGis	PostGIS es un módulo que añade soporte de objetos geográficos a la base de datos objeto-relacional PostgreSQL, convirtiéndola en una base de datos espacial para su utilización en Sistema de Información Geográfica.. [24]	Gratuita
Sqlite	es un sistema de gestión de bases de datos relacional, contenida en una relativamente pequeña (~275 kB) biblioteca escrita en C. SQLite es un proyecto de dominio público creado por D. Richard Hipp. [25]	Gratuita
Spatialite	Es un módulo de extensión que añade soporte a Sqlite para usarla como base de datos de información geográfica. [26]	Gratuita
Google maps	Es un servidor de aplicaciones de mapas en la web que pertenece a Google. Ofrece imágenes de mapas desplazables, así como fotografías por satélite del mundo e incluso la ruta entre diferentes ubicaciones o imágenes a pie de calle Google Street View. [27]	Gratuita

Tabla 2. Herramientas tecnológicas

1.9 MARCO DE REFERENCIA

1.9.1 MARCO TEÓRICO

El hombre siempre ha necesitado el uso de información para tomar decisiones, actualmente, la información se encuentra en grandes volúmenes y muchas veces no es clara y precisa. Es así como surge la necesidad de buscar las técnicas para

filtrar y extraer información que sea exacta, válida y relevante para tomar una decisión o suplir una necesidad.

Hoy en día en pleno apogeo de la globalización este problema surge cada vez con mayor frecuencia cuando se busca información en los diferentes medios. Para ello, se han desarrollado sistemas que aplican en entornos web con el fin de ofrecer servicios personalizados, para evitar la sobrecarga de información y en muchas bibliotecas para darle al usuario recursos de sus preferencias. [28] [29]

A pesar de los muchos proyectos e intentos de satisfacer las cambiantes necesidades del usuario y en consecuencia de las organizaciones; se han generado nuevos requerimientos que pretenden mejorar la forma en que se busca y se filtra la información obteniendo así lo que realmente se necesita y pasando por alto lo que no.

La aplicación de algoritmos de recuperación de información han jugado un papel importante, pues han logrado solucionar varios problemas de este tipo, y ciertamente hoy en día son imprescindibles herramientas para lograr nuestros objetivos de búsquedas, pero no es suficiente para satisfacer las variables exigencias de los criterios de búsqueda de la gran mayoría de usuarios, que buscan de una manera , rápida y fácil, encontrar información y recursos que suplan sus necesidades de información relevante. [29]

Entendemos que tales herramientas (sistemas de recuperación de información) *son imprescindibles pero no suficientes*². A partir de esta necesidad es donde surgen los sistemas de recomendación con el objetivo de facilitar, asistir al usuario en la toma de una decisión, tal decisión en nuestro caso será la selección de opciones sobre establecimientos comerciales que suplirán un requerimiento. Son los sistemas de recomendación los que nos ayudan, asisten, guían, etc. en la toma de decisiones.

Como analogía para facilitar su entendimiento podemos ver el sistema recomendador como un amigo. Usualmente nos relacionamos con personas que comparte nuestros gustos e intereses, si se diera el caso de que requerimos cierta colaboración en una decisión sea de un producto o un servicio acudiremos a nuestros compañeros más cercanos, ellos a través de su experiencia y de nuestro requerimiento nos recomendaran lo que crean pertinente, posteriormente nosotros

²

Vreixo Formoso López, “Técnicas eficientes para la recomendación de productos basadas en filtrado colaborativo.” Universidad de la Coruña, 2014.

tomaremos las mejores recomendaciones y finalmente tomaremos la mejor decisión.

1.9.1.1 SISTEMAS DE RECOMENDACIÓN [6]

1.9.1.1.1 Introducción

Los sistemas de recomendación son herramientas cuyo principal objetivo es el de ayudar a los usuarios en la toma de decisiones (desde el punto de vista de este proyecto de grado tales decisiones serán enfocadas en la selección de un establecimiento. Con el propósito de facilitar el entendimiento nos referiremos a un establecimiento como un ítem) para seleccionar un ítem que más se adecue a su preferencias (Gustos e intereses). [30]

Desde otros puntos de vista definimos los sistemas de recomendación:

Un sistema de recomendación se puede definir, de manera formal, como aquel sistema que tiene como principal tarea seleccionar ciertos objetos de acuerdo a los requerimientos del usuario. Otros autores intentan definir los sistemas de recomendación de una forma más entendible y cercana. Haciendo referencia a éstos como, aquellos sistemas que utilizan las opiniones de los usuarios de una comunidad para ayudar a los usuarios de esa comunidad a encontrar contenidos de su gusto entre un conjunto sobrecargado de posibles elecciones. Estos sistemas son muy atractivos en situaciones donde la cantidad de información que se ofrece al usuario, supera ampliamente cualquier capacidad individual de exploración.³

Teniendo en cuenta lo anterior podemos realizar una clara distinción entre un sistema recomendador y un buscador, en primera instancia el buscador se encargará de filtrar la información solicitada por el usuario a partir de un criterio y por el otro lado el sistema recomendador asiste al usuario a tomar una decisión. Simplificando aún más su definición y con el ánimo de facilitar su entendimiento y “*a riesgo de ser excesivamente simplistas, podemos decir que la tarea de un sistema de recomendación consiste en adivinar qué productos son adecuado para el usuario*”⁴. En este sentido el sistema de recomendación tendrá la capacidad de distinguir si un ítem es adecuado para un usuario basado en sus preferencias particulares.

³ D. L. M. Lope, «*Un nuevo modelo ponderado para Sistemas de Recomendación Basados en Contenido con medidas de contingencia y entro*, España: », Universidad de Jaén, 2012.

⁴Vreixo Formoso López, “Técnicas eficientes para la recomendación de productos basadas en filtrado colaborativo.” Universidad de la Coruña, 2014.

1.9.1.1.2 Fases

En general podemos destacar tres fases de los sistemas de recomendación:

1. Captura de preferencias [6]

Las preferencias reflejan los gustos e interés del usuario, estas serán adquiridas a partir de la interacción del usuario con el sistema; hay dos tipos de preferencias, explícitas e implícitas:

Las explícitas con aquella que el usuario genera de manera consciente, un ejemplo de esta puede ser la valoración de un ítem en un rango de puntuación que refleje su grado de satisfacción o gusto. Las implícitas se infiere a partir de la interacción del usuario con el sistema de manera que el usuario no es consciente, en la sección de [preferencias implícitas y explícitas](#) ahondaremos el tema más detalladamente.

2. Extracción del conocimiento [6]

En esta fase es donde el sistema se encarga de interpretar la información que recopilo en la fase anterior, para poder predecir gustos y preferencias del usuario, en pocas palabras consiste en predecir una puntuación que un usuario habría dado a un ítem. Hay diversos tipos de técnicas y métodos de representación para la extracción del conocimiento que veremos más adelante.

3. Recomendación

A partir del conocimiento adquirido en la fase de extracción del conocimiento, el sistema ya tiene la capacidad de seleccionar los N ítems con mayores puntuaciones que podrían interesarle al usuario, presentándoselos como una lista. En la siguiente figura se resume el funcionamiento de un sistema de recomendación:

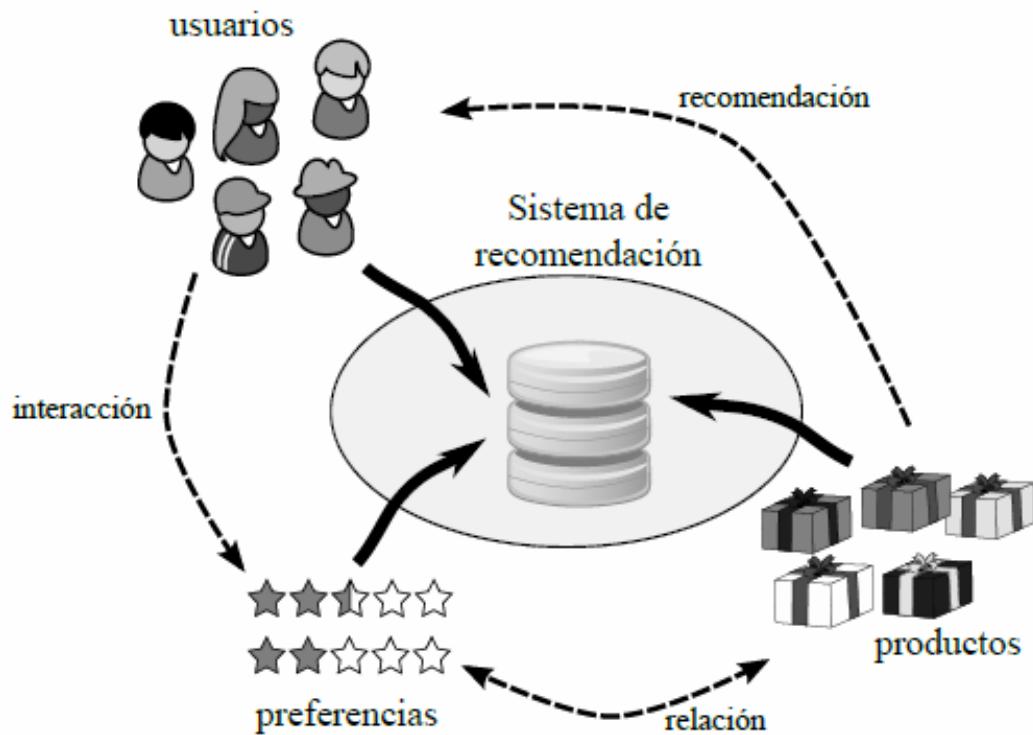


Ilustración 2. Diagrama general de un Sistema de recomendación [6]

1.9.1.1.3 Preferencias implícitas y explícitas [6]

Como ya mencionamos con anterioridad existen dos maneras de capturar las preferencias de los usuarios, la forma explícita e implícita. Es conveniente resaltar que los dos tipos de preferencias son muy diferentes una de la otra y cada una tiene ventajas y desventajas dependiendo del dominio.

En las preferencias explícitas el usuario conscientemente se encarga de valorar los ítems según su gusto e intereses, generalmente se suele usar una escala de puntuación como mecanismo de captura de información. Por otra parte en las preferencias implícitas la opinión del usuario se infiere a partir del uso e interacción con el sistema.

Así, vemos como las preferencias implícitas no tienen la capacidad de reflejar el grado de gusto por parte de un usuario con respecto a un ítem, un gran ejemplo de esto es que si un usuario compra un libro esto no siempre indica que le guste, aumentando aún más la complejidad en la fase de captura, por supuesto también tiene sus ventajas, entre estas están la disponibilidad, pues no dependemos directamente de que el usuario tenga la disposición de dar a conocer sus requerimientos sino que a partir del uso por parte del usuario con la aplicación se

extraen sus preferencias, es decir el sistema no solicite de manera directa al usuario que le proporcione tal información.

Ahora bien, las preferencias explícitas pueden generar un gran impacto en el sistema permitiéndose tener una mayor precisión en la captura de la información, pero dependemos directamente de que el usuario esté dispuesto de transmitirnos esa información y de muchos otros factores que intervienen como por ejemplo el estado de ánimo, los criterios de calificación, etc.

Una de las formas más fáciles de capturar las preferencias explícitamente es mediante el uso de escalas de puntuación, claros ejemplos de esto son el uso de rangos de puntuaciones unarias (Me gusta) que se usa en Facebook, o Google+, un rango de puntuación binario como el de YouTube (Manito arriba o manito abajo), por otro lado encontramos el rango de puntuación de cinco estrellas (Amazon, Netflix,...etc.). Elegir una correcta escala de puntuación es un problema recurrente ya que en la práctica cada tipo de preferencia tiene sus ventajas e inconvenientes, pueden funcionar bien para un dominio pero ser poco práctico para otro [Cosley et al., 2003]. En la siguiente figura encontramos un claro resumen de los diferentes tipos de puntuación y sus características.

	Implícitas	Unarias	Binarias	Escala de puntuaciones
disponibilidad	✓	✓	✓	✓
información aportada	✓	✓	✓	✓
reflejan grado de importancia		✓	✓	✓
reflejan opinión negativa			✓	✓
error y variabilidad	✓	✓	✓	✓

Ilustración 3. Diferentes formas de captura de preferencia [6]

En general el uso de puntuaciones unarias o binarias aunque facilitan su captación no reflejan los distintos grados de satisfacción por parte del usuario produciendo resultados no esperados en ciertos dominios. En este mismo sentido el uso de diferentes niveles nos permite obtener mayor detalle y exactitud en el grado de importancia con la desventaja de aumento de complejidad para el usuario.

“En general, cuanto mayor sea el número de posibles puntuaciones podremos capturar más matices de los productos, pero también estaremos más expuestos a errores debidos a la variabilidad.”⁵

Para este proyecto usaremos la captura de preferencias explícitas con una escala de puntuaciones de cinco estrellas, con el fin de lograr obtener en detalle las preferencias (positivas y negativas) del usuario sin aumentar la complejidad de captura en el portal web.

1.9.1.1.4 Tipos de sistemas de recomendación

Debido al creciente interés por los sistemas de recomendación, en las últimas décadas se crearon y desarrollaron multitudes de técnica que abordan el problema desde diferentes puntos, a continuación veremos de manera general las más conocidas.

- 1 *Sistemas de recomendación Demográficos o basados en información demográfica [31]:*

Los Sistemas de Recomendación Demográficos (SRD) se basan en la idea de que personas con unas características demográficas dadas (edad, sexo, nivel de educación, domicilio) tengan gustos similares a otras personas con características demográficas similares. Aunque en los inicios de los SR se utilizaban mucho los SRD, actualmente se aplican a dominios muy concretos, debido a la dificultad de disponer de los datos demográficos necesarios para su funcionamiento, debido a la preocupación de los usuarios por su privacidad, lo cierto es que su uso es escaso en los sistemas de recomendación ya que no se puede conseguir el nivel de personalización con otras técnicas

- 2 *Sistemas de recomendación basados en contenido [31] [32] [6]*

Un Sistema de Recomendación basado en contenido permite sugerir al usuario nuevos productos en función de su similitud con el contenido (descripción) de otros productos que éste ha juzgado anteriormente, es decir recomienda productos similares que con anterioridad le agradaron al usuario. En la vida cotidiana, a la hora de realizar la descripción de un producto complejo, solemos agrupar su contenido en un conjunto de categorías. La mayor parte de las técnicas basadas en contenido se han utilizado para la recomendación de productos cuyo contenido consiste en texto, como páginas web, noticias, libros, artículos científicos, etc.

⁵ Vreixo Formoso López, “Técnicas eficientes para la recomendación de productos basadas en filtrado colaborativo.” Universidad de la Coruña, 2014.

Se comportan relativamente bien con el contenido basado en texto, pero tiene inconvenientes como su incapacidad para generar recomendaciones inesperadas (los productos son similares) o falta de calidad de las recomendaciones. [6]

3 Sistemas de recomendación basados en conocimiento [31]:

Los Sistemas de Recomendación Basados en Conocimiento son sistemas que utilizan una base de conocimiento que describe cómo los distintos productos satisfacen las necesidades de un usuario y en qué medida (conocimiento funcional del entorno). De esta manera, el sistema encuentra el producto (o productos) que se ajusta a las necesidades que el usuario ha especificado. Esta búsqueda de productos a partir de las necesidades se realiza mediante un proceso de inferencia de algún tipo generalizado como un diálogo entre el sistema y el usuario en el que el sistema al usuario sobre ciertos aspectos y a partir de esa información se obtiene los productos con tales características.

Además de tener información sobre la características de los productos, necesitan información sobre las relaciones entre los mismos, su capacidad para satisfacer las necesidades del usuario, etc. Esto puede resultar una tarea costosa, ya que se necesita realizar un proceso de Ingeniería del Conocimiento para obtenerla.

En general este tipo de sistemas de recomendación son realmente útiles en dominios donde resulta complejo o es difícil adquirir las preferencias de los usuarios o donde el costo de la recomendación es muy alto. Por otro lado este sistema de recomendación no es netamente automático ya que se necesitan de un grupo de expertos del dominio que modele el conocimiento y por el lado del usuario se necesita un alto grado de interacción con el sistema.

4 Sistemas de recomendación de filtrado colaborativos [31]:

Estos funcionan basándose en el concepto del “boca a boca” (en inglés, “word-of-mouth”), que calcula recomendaciones para un usuario concreto utilizando la información proporcionada por el resto de los usuarios que hay en el sistema., están basados enteramente en opinión de los propios usuarios, dando muy buenos resultados.

Tienen la ventaja de que no es necesario tener información sobre los productos que se van a recomendar. Esto es porque estos sistemas tratan los productos como una caja negra, de la que sólo se conocen las valoraciones de preferencia que los distintos usuarios del sistema han dado sobre los mismos. Otra ventaja de los SRC es que su precisión mejora con el tiempo, cuantos más usuarios haya en el mismo y mayor cantidad de valoraciones compartan con el resto, mejores serán sus resultados.

Los sistemas de recomendación basados en filtrado colaborativo tienen ciertas desventajas que se han intentado solucionar: necesitan una gran cantidad de datos para funcionar, sufren problemas escalabilidad.

En el segmento posterior detallaremos más los sistemas de recomendación basados en filtrado colaborativo.

5 *Sistemas de recomendación híbridos* [33]:

Investigaciones recientes han demostrado que un enfoque híbrido que combina filtrado colaborativo y el filtrado basado en contenido puede ser más eficaz en algunos casos. Los enfoques híbridos se pueden implementar de varias maneras:

- 1 haciendo predicciones basadas en el contenido y las recomendaciones basadas en colaboración por separado para luego combinarlos, agregando nuevas capacidades.
- 2 unificación de los criterios en un solo modelo.

No obstante su complejidad de desarrollo y mantenimiento puede aumentar.

Para nuestro proyecto usaremos el filtrado colaborativo como método principal de recomendación debido a sus grandes ventajas sobre las otras técnicas, por consiguiente detallaremos el tema del filtrado colaborativo en la siguiente sección.

1.9.1.1.5 *Filtrado colaborativo*

Como ya vimos con anterioridad el filtrado colaborativo es una técnica que gano popularidad gracias a los buenos resultados que se obtenían, al estar basada enteramente por las preferencias de otros usuarios con similares gustos aumenta la calidad de la recomendación siendo este proceso muy similar al de la vida cotidiana en donde generalmente transmitimos nuestro conocimiento a otras persona con la esperanza de que les sea de utilidad y solemos ignorar aquellas recomendaciones u opiniones de personas no lo suficiente similares a nosotros, es decir confiamos en aquellas personas que tienen gustos e intereses similares e ignaros las que no.

Esta es, precisamente, la forma en que funcionan los sistemas de filtrado colaborativo. La idea es que si dos usuarios tienen un patrón de preferencias similar, se debe a que comparten gustos o intereses, y por tanto los productos que han sido de utilidad a uno de ellos pueden ser una buena recomendación para el otro usuario.

El papel del algoritmo se centra en buscar relaciones por lo que puede ser abordado desde diferentes técnicas, tales técnicas en la literatura se dividen las basadas en memoria y las basadas en modelos. Las basadas en memoria procesan la matriz de puntuaciones en cada recomendación, mientras que las basadas en modelos utilizan algoritmos de aprendizaje automático para construir un modelo. En esta

sección veremos las diferentes técnicas del filtrado colaborativo así como sus ventajas y problemas.

1.9.1.1.5.1 Ventajas y limitaciones del filtrado colaborativo [6]

Entre las principales ventajas del filtrado colaborativo tenemos:

- Independencias usuarios vs productos:
Desde el punto e vista del sistema los usuarios y productos son vistos como identificadores asociados a un conjunto de preferencias, por lo tanto no necesitamos tener un conocimiento avanzado de las características y naturaleza de dominios, es decir podemos aplicar el filtrado colaborativo para multitud de dominios diferentes.
- Calidad evaluada por personas:
Al ser los usuarios quienes evalúan los ítems nos dan una gran ventaja en la calidad de las recomendaciones permitiendo que estas sean más precisas.
- Recomendación de nuevos productos o inesperados:
Como el filtro colaborativo se basa enteramente en las preferencias tiene la capacidad de recomendar ítems inesperados que no guardan una relación aparente con el contenido del ítem. Este tipo de recomendaciones son muy valiosas en los sistemas de recomendación ya que es posible que el usuario ni siquiera habría conocido el ítem si no fuera por el sistema de recomendación.

Claro está que el filtrado colaborativo también está sujeto a limitaciones, entre estas tenemos:

- Escasez de puntuaciones:
Es muy común que las puntuaciones que otorga un usuario sean muy escasas con respecto a la cantidad de ítems que hay en el sistema, esto generalmente es causado por que el usuario no tiene total conocimiento de todos los productos sino que ha experimentado un pequeño conjunto de estos. Esto por supuesto es un grave problema conocido como *sparcility problema* que causa una baja densidad en la matriz de votaciones afectando encontrar la similitud de los usuarios por consiguiente la propia recomendación.
- *Cold-start* o problema de arranque en frío:
Está muy relacionado con el problema anterior, y es que cuando se agrega un usuario nuevo o un ítem nuevo, en el caso de los usuarios, todavía no habrán puntuado suficientes productos para recibir recomendaciones de calidad, para el caso de los ítems esto no habrán sido suficientemente puntuados para poder ser recomendados.
- Ataques y manipulaciones de preferencias:

Al ser una técnica enteramente basada en las preferencias que proporcionan los usuarios es susceptible a los ataques de *spam*.

- Privacidad:
La captura de información del usuario por parte del sistema puede ser interpretada como un problema de privacidad.
- Escalabilidad:
Las técnicas de filtrado colaborativo son especialmente sensibles a la escalabilidad y eficiencia, en general las técnicas más complejas o que tengan en cuenta un número elevado de preferencias serán propensas a estos problemas conforme se agregan usuarios nuevos y productos en el tiempo.

1.9.1.1.5.2 Principales técnicas del filtrado colaborativo [6]

En la literatura es común encontrar que las técnicas de filtrado colaborativo se encuentran clasificadas en algoritmos basados en memoria y basados en modelos de acuerdo a la manera en que se procesa la información a partir de la matriz de puntuaciones.

Así los algoritmos basados en memoria procesan la matriz cada vez que calculan una predicción o recomendación, generalmente usan medidas de similitud para encontrar en los usuarios o productos un patrón de similitud que son usados para realizar la recomendación. Por otro lado los algoritmos basados en modelo utilizan la matriz de votos para entrenar un modelo creado con anterioridad, el cual intenta evidenciar relaciones o características entre ítems y usuarios, estas predicciones se realizan directamente a partir del modelo si realizar un nuevo procesamiento de la matriz, solo se actualizarán cada cierto tiempo o cuando los datos de puntuación varíen.

A continuación veremos de manera general las principales técnicas del filtrado colaborativo:

1.9.1.1.5.3 Algoritmos basados en vecinos

Estos algoritmos son unas de las técnicas más populares y conocidas entre los sistemas de recomendación basada en filtrado colaborativo, se centra en buscar los vecinos ya sean de los usuarios o los productos (usuarios o productos con preferencias similares) y a partir de los resultados generar las recomendaciones.

Usualmente suelen dividir el funcionamiento del algoritmo en tres fases principales:

- Calculo de similitud entre los usuarios
- Selección de vecindario según la similitud previamente calculada

- Calculo de la predicción o recomendación a partir de las puntuaciones de los vecinos

Algoritmos basados en usuarios

Como su nombre lo indica, selecciona el vecindario basado en los usuarios similares al usuario actual, su similitud es calculada comparando las puntuaciones de los ítems que ambos usuarios han puntuado, si un usuario tiene puntuaciones similares al de otros usuarios se asume que el resto de puntuaciones también será parecida generando una recomendación de aquellos productos bien puntuados por sus vecinos. En la siguiente imagen se ilustra el enfoque:

		productos				
		Gift 1	Gift 2	Gift 3	Gift 4	Gift 5
usuarios	User 1	★★★		★★★		★★★
	User 2	★★★		★★★		
	User 3		★★★	★★★		★★★
	User 4	★★★			★★★	
	User 5		★★★	★★★		★★★
	User 6					

Ilustración 4. Algoritmos basados en usuarios. [6]

En la literatura se encuentran distintas técnicas para encontrar medidas de similitud, entre estas tenemos:

- Coeficiente de correlación de Pearson
- Coeficiente de Pearson forzado
- Coseno
- Coeficiente de Pearson ponderado
- Diferencia cuadrática media

Una vez calculada la similitud de los usuarios se seleccionan como vecinos aquellos más semejantes al actual, entre las técnicas más conocidas tenemos:

- Umbral de correlación

- Vecinos similares

Así mismo en la literatura los algoritmos basados en usuarios se han utilizado fuertemente para la tarea de predicción, donde hay varias alternativas basadas en las puntuaciones otorgadas por los vecinos, las más conocidas son:

- Media ponderada por similitud
- Media normalizada
- Clasificación

Algoritmos basados en productos

Estos algoritmos son similares a los algoritmos basados en usuarios descritos anteriormente, su diferencia radica en que en qué lugar de buscar usuarios similares se basan en buscar productos similares a los puntuados, así su enfoque se representa en la siguiente figura:

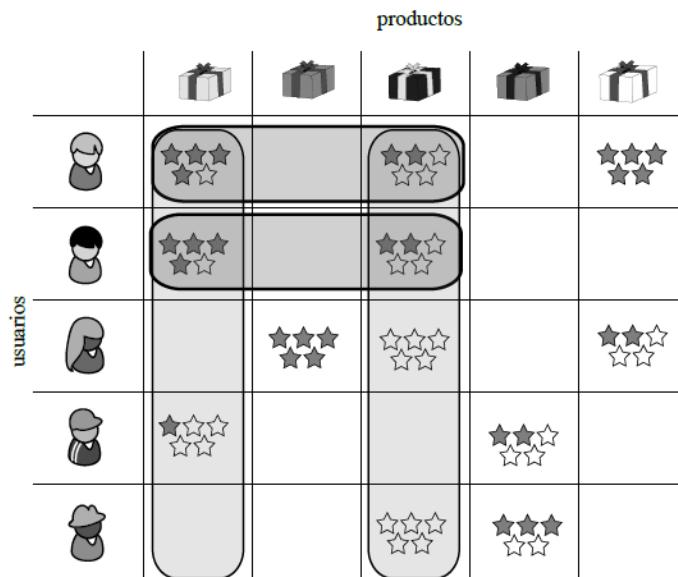


Ilustración 5. Algoritmos basados en productos

Entre las medidas de solicitud más conocidas son:

- Coseno
- Coseno ajustado
- Basado en probabilidad condicional

En la tarea de predicción los algoritmos son similares a los tratados en los algoritmos basados en usuarios.

Similarity fusión

Es un algoritmo similar a los anteriores, pero en lugar de tener en cuenta el enfoque basado en usuarios o productos por aparte tiene en cuenta ambos, con el objetivo de aprovechar la mayor cantidad de información y mejorar los resultados en situaciones de baja densidad.

Algoritmos basados en regresión

Este tipo de algoritmos están basados enteramente en la tarea de predicción, usando la puntuación de productos para la estimación. Para realizar esto el algoritmo entrena un modelo de regresión para cada par de productos utilizando los datos disponibles en la matriz de puntuaciones, es decir que tendrá la capacidad de estimar una calificación de un producto a partir de otro. Se parte de que un usuario habrá calificado varios productos por lo que las estimaciones obtenidas son combinadas para obtener la predicción final de la siguiente forma:

$$\hat{r}_{ai} = \sum_{j \in J^{(a)}}^* f_{ji}(r_{aj})$$

Regresión lineal

Este algoritmo utiliza un modelo de regresión lineal que asume que existe una relación lineal entre las puntuaciones en el que en la fase de entrenamiento se estiman los parámetros del modelo a partir de los datos en la matriz de votos.

Slope one

Los algoritmos de *Slope one* es un tipo de algoritmo enfocado en la tarea de predicción donde asumen que la pendiente de la recta de regresión es igual a 1, usando la diferencia media entre las puntuaciones de ambos productos para el cálculo del parámetro indicado que es entonces una diferencia de promedio entre la valoración de dos elementos. Esto ha demostrado ser mucho más exacto que la regresión lineal en algunas instancias, y toma la mitad del espacio de almacenamiento o menos.

Algoritmos de clasificación de la puntuación

Al igual que el algoritmo anterior, este se enfoca en la tarea de predicción, pero abordando el problema desde la perspectiva de la clasificación, donde cada uno de los posibles valores de la matriz de votación es vista como una clase, por lo que se intenta predecir a cuál de las posibles clases pertenece realmente la puntuación dada por el usuario. Entre las técnicas conocidas tenemos:

- Naive Bayes
- Personality Diagnosis

Agrupamiento o clustering

Estos tipos de algoritmos consisten en la agrupación ya sea de usuarios y/o ítems de acuerdo a sus puntuaciones, esto debido a que asumen que hay grupos de usuarios e ítems con características similares, por lo tanto una vez que se asigna a un grupo la predicción es obtenida por las puntuaciones otorgadas por los usuarios del mismo grupo.

Existen varias técnicas de agrupamiento entre las más conocidas están *k-means*, donde se usa un proceso iterativo que minimiza la distancia de los usuarios con el perfil medio del grupo al que pertenecen, otras técnicas usan modelos más completos como *co-clustering*, donde se agrupan usuarios y productos grupos superpuestos.

Por otro lado es necesario resaltar que estos algoritmos suelen ser usados en combinación con otros algoritmos con los objetivos de incrementar la matriz de puntuaciones y de eficiencia sustituyendo a los usuarios en ciertas fases del algoritmo reduciendo el volumen a manejar.

Modelos probabilísticos

De manera muy similar a los anteriores, esta técnica se basa en la existencia de clases que determinan usuarios o productos median el uso de modelos probabilísticos donde la pertenencia a una clase es modelada como una variable latente, así la clase a la que pertenece ya sea un usuario o un producto determinara la distribución de las puntuaciones. De igual manera que los algoritmos de clasificación de puntuación el modelo se usa para estimar la probabilidad de que un usuario puntué a un producto de cierta forma.

Entre las técnicas más populares tenemos:

- Agrupamiento basado en un modelo bayesiano
- Modelo de aspectos
- Modelos mixtos

Técnicas de reducción de la dimensionalidad de la matriz

Es común encontrar que en la realidad la matriz de votaciones sea muy grande y dispersa. Es por esto que las técnicas de reducción de la dimensionalidad tienen como objetivo transformar esta matriz en una de menores dimensiones que refleje las características de la matriz original a la hora de realizar la predicción, una vez seleccionadas la información se representará con una matriz de reducidas dimensiones que represente el grado de afinidad de un usuario con las

características y el grado de afinidad de un ítem a las mismas. Así la idea es que dichas características representen factores latentes en la matriz original.

Esto permite acelerar el proceso de recomendación ya que solo habrá que considerar las características en lugar de todas las votaciones. Por otro lado minimiza los problemas relacionados con la dispersión de la matriz y la presencia de datos erróneos.

Entre las técnicas más conocidas tenemos

- SVD (Descomposición singular de valores)
- RSVD (SVD mejorado)
- NSVD2
- SVD++

Como ya se mencionó con anterioridad, nuestro proyecto de grado hará uso de las técnicas de reducción de dimensionalidad de SVD, por lo tanto trataremos este tema en detalle en las siguientes secciones.

Modelo integrado

Cada una de las técnicas que se presentaron con anterioridad tiene sus ventajas y limitaciones al igual que su enfoque en ciertos dominios. El diseño de modelos integrados permite aproximar de manera más directa los resultados obviando debilidades propias de cada tenía.

EL problema podría ser en que las técnicas serían más específicas y dependientes de un dominio de datos, al igual que el aumento de su complejidad, lo que no siempre es lo deseable.

1.9.1.2 Descomposición en Valores Singulares (SVD) [7]

La descomposición singular de valores (*Singular Value Decomposition en adelante SVD*) es una técnica de factorización de matrices que permite descomponer una matriz A en otras matrices U, S, V de la siguiente forma:

$$SVD(A) = U \times S \times V^t$$

Es decir el producto matricial de U por S por V^t da como resultado la matriz A.

Con respecto a las dimensiones partimos de que la matriz A tendrá unas dimensiones de $n \times m$ donde n representa las filas y m representa las columnas. La matriz U será una matriz ortogonal que tendrá dimensiones de $n \times n$, y la matriz V de $m \times m$ también ortogonal, así tenemos que las matrices U y V cumplirán las siguientes propiedades:

$$U^t \times U = U \times U^t = I_{n \times n}$$

$$V^t \times V = V \times V^t = I_{m \times m}$$

La matriz S será una matriz diagonal de dimensiones de $m \times n$ la cual tendrá en su diagonal lo que se denomina valores singulares puesto en orden decreciente siendo valores mayores o iguales a cero, es decir:

$$S = \begin{bmatrix} \lambda_1 & 0 & \dots & 0 \\ 0 & \lambda_2 & 0 & \vdots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \lambda_n \end{bmatrix}$$

Por lo tanto la descomposición será igual a:

$$SVD(A)_{n \times m} = U_{n \times n} \times S_{n \times m} \times V^t_{m \times m}$$

Aunque no detallaremos el proceso de factorización del SVD si destacaremos una propiedad bastante interesante que se aplicaba especialmente bien no solo en los sistemas de recomendación sino también para otras áreas. Esta consiste en que rediciendo el número de valores singulares de la matriz S a los primeros k valores, se obtendrá una aproximación de la matriz original A, que permite ser reconstruida a partir de las versiones reducidas de las otras matrices cometiendo un cierto error, pero disminuyendo el tamaño. Es decir:

$$SVD(A)_{n \times m} \cong U_{n \times k} \times S_{k \times k} \times V^t_{k \times m}$$

Esta importante propiedad es derivada del Teorema de *Eckart-Young* que nos dice que la mejor aproximación a la matriz original A, la obtendremos poniendo a 0 los n valores singulares más pequeños, así se reducirán las matrices al número de valores singulares no nulos que tenga la matriz S.

Esto resulta entonces en la transformación de gran cantidad de datos en la representación reducida de los mismo, siendo por lo tanto una propiedad muy importante nos da la capacidad de reducir considerablemente el tiempo de computo de cálculo y de uso de memoria para las tres matrices .

1.9.1.3 SVD aplicado al filtrado colaborativo [7]

Ya hablamos sobre los sistemas de recomendación, el filtrado colaborativo y sobre el SVD, en esta sección detallaremos el enfoque del SVD aplicado al filtrado colaborativo y como este puede llegar a realizar muy buenas recomendaciones en comparaciones con otras técnicas, pero con los problemas de escalabilidad y eficiencia.

Dentro del filtrado colaborativo, como ya vimos es necesario el manejo de una matriz de votaciones de usuarios e ítems, así podemos considerar esta matriz como la base para aplicar el SVD y obtener una factorización con las matrices U, S, V. Luego podremos reducir sus dimensiones con el teorema de *Eckart-Young* a k dimensiones así nuestra matriz factorizada sería:

$$SVD(A)_{\text{usuarios} \times \text{items}} \cong U_{\text{usuarios} \times k} \times S_{k \times k} \times V^t_{k \times \text{items}}$$

Donde A es la matriz de votaciones.

Por otro lado podemos simplificar el proceso de SVD obteniendo solamente dos matrices que representen los factores de usuarios e ítems, esto es posible descomponiendo la matriz S en dos matrices iguales (haciendo multiplicación de raíz cuadrada) y multiplicando cada una de ellas con la matriz U y V. Por lo tanto obtendremos las matrices de factores de usuarios *Ufac* y factores de ítems *Ifac* como lo veremos a continuación:

$$S_{k \times k} = \begin{bmatrix} \lambda_1 & & \\ & \lambda_2 & \\ & & \lambda_n \end{bmatrix} = \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix} x \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

$$\begin{aligned} & [Matrix\ de\ votaciones]_{n \times m} \\ &= [U_{n \times k}]x \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix} x \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix} x [V_{k \times m}] \end{aligned}$$

$$\begin{bmatrix} & Ufact & \\ & & \end{bmatrix} = [U_{n \times k}]x \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

$$\begin{bmatrix} & Ifact & \\ & & \end{bmatrix} = [V_{k \times m}]x \begin{bmatrix} \sqrt{\lambda_1} & & \\ & \sqrt{\lambda_2} & \\ & & \sqrt{\lambda_n} \end{bmatrix}$$

Por lo tanto

$$[Votos_{n \times m}] = [Ufac] x [Ifac]$$

Ahora bien ¿Qué representan los factores de usuarios y los de ítems? La verdad es que no existe ninguna explicación sobre el significado de estos factores, tan solo es posible decir que estos representan de alguna manera características de usuarios e ítems e incluso esta puede ser real (edad, sexo,..etc.) Pero SVD no nos dice que característica está asociada cada factor, pero aun así tal característica nos permite predecir ítems que no han sido votados por los usuarios.

Este proceso no es obvio y en la mayoría de casos es muy difícil e imposible encontrar su significado, pero nos permite distinguir la similitud que hay entre usuarios e ítems, calculando la distancia entre ellos a partir de sus factores.

1.9.1.3.1 Problemas del SVD en el filtrado colaborativo

Dentro de las técnicas del SVD aplicadas al filtrado colaborativo se presentan problemas, como la forma en que se tratan aquellos ítems que el usuario no ha valorado, ya que si este ítem es valorado con cero la predicción será de valores próximos a cero, que no es un resultado esperado ya que se pretende obtener la predicción de la puntuación otorgada a ese ítem por parte del usuario, de igual forma el otro problema planteado es el de la escasez de puntuaciones en la matriz que ya comentamos con anterioridad.

Para entender a fondo este problema veremos un breve ejemplo:

Supongamos que tenemos la siguiente matriz de votaciones, donde las filas representan a los usuarios y las columnas los ítems:

Usuarios	Ítems				
	I1	I2	I3	I4	I5
U1	1	3	5	3	1
U2	1	4	4	3	2
U3	1	?	?	3	?

Si cambiamos los valores no conocidos por ceros tendremos:

Usuarios	Ítems				
	I1	I2	I3	I4	I5
U1	1	3	5	3	1
U2	1	4	4	3	2
U3	1	0	0	3	0

Ahora al aplicar el SVD con reducción de k=2 obtenemos lo siguiente:

$$\begin{bmatrix} 1 & 3 & 5 & 3 & 1 \\ 1 & 4 & 4 & 3 & 2 \\ 1 & 0 & 0 & 3 & 0 \end{bmatrix} \approx \begin{bmatrix} -0,69 & 0,11 \\ -0,70 & 0,13 \\ -0,17 & -0,98 \end{bmatrix} \cdot \begin{bmatrix} 9,58 & 0 \\ 0 & 2,75 \end{bmatrix} \cdot \begin{bmatrix} -0,16 & -0,51 & -0,65 & -0,49 & -0,22 \\ -0,27 & 0,31 & 0,38 & -0,81 & 0,13 \end{bmatrix}$$

Ahora al tener ya las matrices U, S, V podremos obtener la matriz original, al hacer el producto matricial obtendremos:

$$\begin{bmatrix} 1 & 3,47 & 4,46 & 3 & 1,49 \\ 0,99 & 3,53 & 4,53 & 2,99 & 1,52 \\ 0,99 & -0,01 & 0,01 & 2,99 & -0,01 \end{bmatrix}$$

Si bien los valores conocidos de la matriz original son muy bien aproximados tenemos un problema con los valores no conocidos, ya que estos los aproxima a

cero, debido a este inconveniente es donde surge la propuesta de Simon Funk que explicaremos más adelante.

1.9.1.3.2 RSVD *Simon Funk*

En el contexto del concurso organizado por Netflix Simon Funk propuso un nuevo método para aproximar dicha factorización, solucionando problemas de dispersión y escalabilidad, tal método acabó teniendo un gran éxito en el filtrado colaborativo.

Lo que tenemos que buscar son los factores tanto de usuarios e ítems a partir de la matriz de votación, este problema por lo tanto puede abordarse a partir de problemas de regresión, el problema entonces queda en encontrar los valores de las matrices de factores que satisfagan la siguiente aproximación

$$\begin{bmatrix} A_{1,1} & \cdots & A_{1,m} \\ \vdots & \ddots & \vdots \\ A_{1,m} & \cdots & A_{n,m} \end{bmatrix} = \begin{bmatrix} u_{1,1} & \cdots & u_{1,k} \\ \vdots & \ddots & \vdots \\ u_{n,1} & \cdots & u_{n,k} \end{bmatrix} \times \begin{bmatrix} i_{1,1} & \cdots & i_{1,m} \\ \vdots & \ddots & \vdots \\ i_{k,1} & \cdots & i_{k,m} \end{bmatrix}$$

La diferencia radica en que en este caso no tendremos en cuenta la matriz de votos, cada voto será el producto escalar de los factores de usuarios y los factores de los ítems, así si consideramos a q_i al vector que represente los factores de un ítem y p_j al vector que representa los factores del usuario se tendría que cumplir:

$$A_{i,j} = q_i^t \cdot p_j \cong \sum_{k=1}^k q_{ik} p_{kj}$$

Ahora solo se ajustan los factores de aquellos usuarios que han definido un voto reduciendo considerablemente el sistema de ecuaciones y solucionando el problema de la dispersión, cometiendo in breve error en los valores iniciales que es algo natural de SVD en el filtrado colaborativo y que gracias a ello se pueden obtener las predicciones de los ítems no votados por un usuario. Tal error lo podemos definir de la siguiente manera:

$$e_{i,j} = |A_{i,j} - q_i^t \cdot p_j|$$

Debido a que el error absoluto es una función complicada de tratar, usaremos el error cuadrado medio, así tenemos:

$$(e_{i,j})^2 = (A_{i,j} - q_i^t \cdot p_j)^2 \cong (A_{i,j} - \sum_{k=1}^k (p_{ik} \cdot q_{kj}))^2$$

Por lo tanto lo que plante Simon Funk es encontrar el error mínimo para hacer la mejor aproximación de la siguiente forma:

$$\text{Minimize}_{p^*, q^*} = \sum (A_{i,j} - q_i^t \cdot p_j)^2$$

Es decir aquella que minimice todos los errores que se pueden cometer al calcular los factores de los usuarios y de los ítems.

Para encontrar el mínimo de la función ya detallada existen como la del *descenso del gradiente estocástico* (SGD stochastic gradient descent), o mínimo de cuadrados alternantes ALS, que lo que buscan es ajustar las matrices de factores poco a poco de manera automática.

1.9.1.3.3 Regularización u overfitting

Conociendo el objetivo del SVD aplicado al filtrado colaborativo que es encontrar el mínimo del error, es necesario mencionar un problema asociado a los algoritmos de aprendizaje conocido en la literatura como *overfitting* o sobre especialización. Este problema se genera al momento de ajustar los datos de entrenamiento al realizar la regresión, ajustándose muy bien pero de igual forma también se ajustaran muy bien a los datos no puntuados o cercanos a cero, en la siguiente figura es un ejemplo claro de *overfittin*:

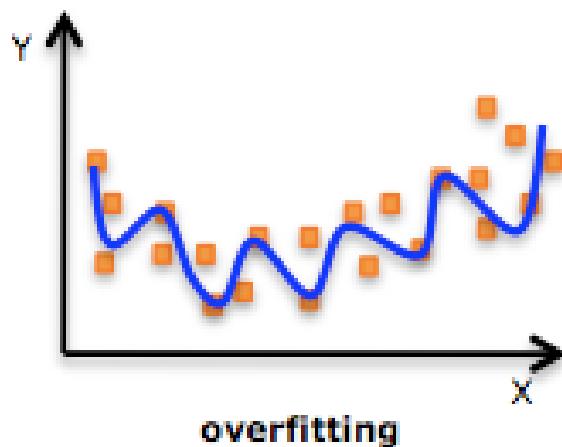


Tabla 3. Problemas de Regularización

Como podemos ver, la función se ajusta a los puntos conocidos y por lo tanto no hay una generalización, en este caso se manejó para dos dimensiones pero en el caso del SVD es para n dimensiones.

Para evitar entonces este tipo de problemas lo que hacemos es agregar una constante de regularización al problema de encontrar el mínimo del error, esto nos da la ventaja de que la regresión que se realiza no se ajuste tan detalladamente a los datos de entrenamiento, es decir a los votos conocidos y por lo tanto nos permite generalizar más los factores de usuario y los factores de ítems, esto se verá reflejando al momento de realizar la recomendación. Nuestra nueva expresión sería:

$$\text{Minimize}_{p^*, q^*} = \sum (A_{i,j} - q_i^t \cdot p_j)^2 + \frac{\lambda}{2} (\|q_i^2\| + \|p_j^2\|) \quad (1)$$

Donde λ es una constante de regularización

Si representamos lo anterior en una función de dos dimensiones obtendríamos esto:

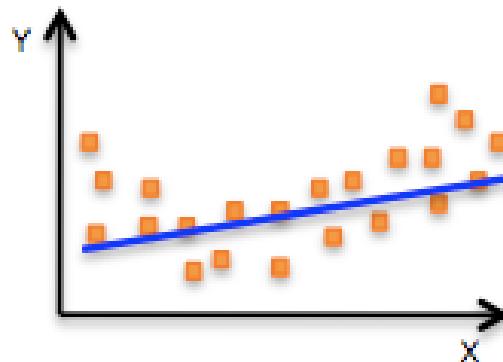
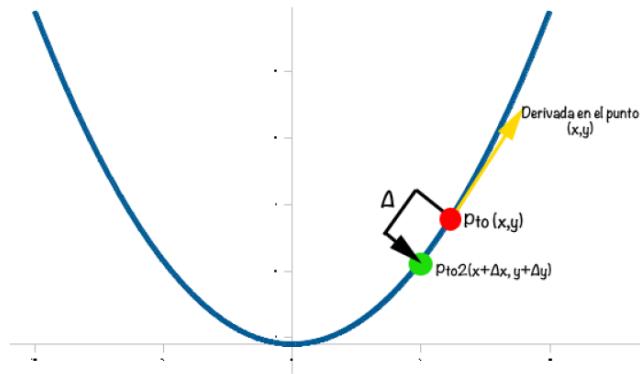


Tabla 4. Regularización

1.9.1.3.4 Descenso del gradiente estocástico (SGD)

Teniendo claro el objetivo principal del SVD para el filtrado colaborativo planteado por es necesario encontrar el mínimo de la función (1) o acercarnos lo suficiente a la solución. Como ya se comentó con anterioridad existen dos aproximaciones para poder hallar esta solución, ya sea mediante el SGD o ALS. Para nuestro proyecto usaremos la técnica del SGD debido a su simplicidad al momento de implementar y a qué el solo se contará con un equipo para su procesamiento.

La técnica del SGD consiste en ir moviéndose poco a poco por la función (de n dimensiones) hasta encontrar el mínimo, nos moveremos evaluando la función en un punto dado y moverse una distancia hacia otro punto, ahora se calcula la derivada en ese punto y nos moveremos al lado opuesto de la derivada de la función que se quiere minimizar, en dos dimensiones un ejemplo se vería como este:



Como podemos observar la función se va moviendo una cierta distancia Δ hacia el mínimo de la función, que se encuentra al sentido opuesto de la derivada evaluada en el punto inicial, por otro lado también hay que tener en cuenta el número de desplazamientos que realizaremos a partir del punto inicial, este valor comúnmente es llamado *Ephocs*.

Ahora bien este método tiene sus inconvenientes, entre esto están el número de veces que tenemos que realizar el desplazamiento y la distancia que se desplazará.

Por otro lado tampoco sabremos si el mínimo que nos da SGD es realmente el mínimo o un punto muy cercano al mínimo, pero en realidad esto no es un problema muy importante en el enfoque de los sistemas de recomendación (a menos de que sea un mínimo local) ya que lo que pretendemos es encontrar una buena solución o una muy cercana a esta.

Ahora bien ya conociendo la técnica del SGD podemos aplicarla directamente al nuestra expresión del error cuadrático derivando con respecto a p_{ik} y q_{kj} de la siguiente manera:

$$(e_{i,j})^2 = (A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj})^2 + \frac{\lambda}{2} (q_i^2 + p_j^2)$$

$$\frac{\partial(e_{i,j})^2}{\partial p_{ik}} = 2 \left(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj} \right) (-q_{kj}) + \frac{\lambda}{2} (2 p_{ik})$$

$$\frac{\partial(e_{i,j})^2}{\partial q_{kj}} = 2 \left(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj} \right) (-p_{ik}) + \frac{\lambda}{2} (2 q_{kj})$$

A partir de lo anterior tenemos

$$\frac{\partial(e_{i,j})^2}{\partial p_{ik}} = -2(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj})(q_{kj}) + \lambda(p_{ik}) \quad (2)$$

$$\frac{\partial(e_{i,j})^2}{\partial q_{kj}} = -2(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj})(p_{ik}) + \lambda(q_{kj}) \quad (3)$$

Teniendo ya el gradiente nos queda aplicar las reglas de actualización tanto para p_{ik} como para q_{kj} (2) (3) con las constantes de regularización y aprendizaje. Así nuestra regla de actualización parte de $\sigma_{n+1} = \sigma_n - \alpha \nabla f(x)$, donde σ_{n+1} es nuestro nuevo valor en la matriz de votaciones, σ_n su valor actual, α constante de aprendizaje y $\nabla f(x)$ es el gradiente ya obtenido. Por lo tanto nos queda que:

Para p_{ik}

$$p'_{ik} = p_{ik} - \alpha (-2 \left(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj} \right) (q_{kj}) + \lambda(p_{ik}))$$

$$p'_{ik} = p_{ik} + \alpha (2(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj})(q_{kj}) - \lambda(p_{ik})) \quad (4)$$

Para q_{kj}

$$q'_{kj} = q_{kj} - \alpha \left(2 \left(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj} \right) (p_{ik}) + \lambda(q_{kj}) \right)$$

$$q'_{kj} = q_{kj} - \alpha + (2(A_{i,j} - \sum_{k=1}^k p_{ik} \cdot q_{kj})(p_{ik}) - \lambda(q_{kj})) \quad (5)$$

Nuestro enfoque se plasma en las expresiones (4) y (5) que serán las que darán quienes actualizaran los valores de la matriz de votaciones a partir de cada *Ephoch*, tanto para los factores de usuarios, como para los factores de ítems.

1.9.1.3.5 Pseudocódigo

Partiendo de lo anterior ya tenemos la capacidad de plasmar las técnicas del SVD enfocado al filtrado colaborativo, presentamos entonces el pseudocódigo realizado en el lenguaje de programación Python:

```

9  @INPUT:
10     R      : la matrix que se factorizara de dimenciones N x M
11     P      : una matriz inicializada con valores aleatorias, su tamano es de N x K
12     Q      : una matriz inicializada con valores aleatorias, su tamano es de M x K
13     K      : el numero de factores latentes
14     EPOCHS : el numero maximo de EPOCHS
15     ALPHA   : taza de apredizaje
16     LAMBDA  : parametro de regularizacion
17 @OUTPUT:
18     Matrices finales P y Q
19 """
20 def matrix_factorization(R, P, Q, K, EPOCHS=5000, ALPHA=0.0002, LAMBDA=0.02):
21     Q = Q.T
22     ass=1
23     for step in xrange(EPOCHS):
24         for i in xrange(len(R)):
25             for j in xrange(len(R[i])):
26                 if R[i][j] > 0:
27
28                     eij = R[i][j] - numpy.dot(P[i,:],Q[:,j])
29                     for k in xrange(K):
30                         P[i][k] = P[i][k] + ALPHA * (2 * eij * Q[k][j] - LAMBDA * P[i][k])
31                         Q[k][j] = Q[k][j] + ALPHA * (2 * eij * P[i][k] - LAMBDA * Q[k][j])
32     return P, Q.T

```

Tabla 5. Pseudocódigo de RSVD

Es necesario que antes de pasar las matrices Q y P (Usuarios e ítems) rellenarlas con valores aleatorios. Por otro lado es necesario decir que se usó la librería de numpy⁶ para las multiplicaciones de matrices usando la función `dot`.

Como se puede ver en el algoritmo se va ajustando las matrices de factores de usuarios y factores de ítems, factor a factor, así en la primera tarea se modificará el primer factor del primer usuario, que haya votado el primer ítem, luego se ajustará el primer factor del usuario y el primer factor del segundo ítem votado, y así respectivamente hasta completar los factores de usuarios y los factores de ítems.

Resumiendo, por cada *Ephochs* se ajustaran dos factores, uno de usuarios y otro de los ítems, esto solo se realiza en los votos emitidos por los usuarios.

Como bien podemos ver, el algoritmo tiene una complejidad muy alta, para nuestro proyecto de grado usaremos la implementación PyRecsys⁷ creado por Oscar Celma el cual tiene la implementación del SVD y con ciertas mejoras.

1.9.1.4 Portal web

Un portal es un término asociado a puerto, utilizado para referirse a un punto de partida o de inicio en Internet, en el que se organizan sus contenidos, ayudando al usuario en la actividad de navegación y concentrando servicios y productos que son atractivos para él, de forma que le permita realizar cuanto necesite hacer sin salir del sitio. [34]

⁶ <http://docs.scipy.org/>

⁷ Oscar Celma

Las características fundamentales con que se puede identificar un portal frente a otro tipo de páginas web son según David Morrison⁸, utilizando las iniciales del término son las siguientes:

- Personalización para usuarios finales.
- Organización del escritorio.
- Recursos informativos divididos y organizados.
- Trayectoria o seguimiento de las actividades de los usuarios (tracking).
- Acceso a bases de datos.
- Localización de gente o de cosas importantes. [35]

En conclusión, un portal web es un sitio que permite a un usuario concentrar su navegación en un solo punto y encontrar en este lo que para él es atractivo, personalizando la información y adaptándose a las necesidades del usuario.

1.9.2 MARCO HISTÓRICO

El grupo de investigación METIS de la Universidad Distrital se ha preocupado siempre por aplicar tecnologías innovadoras en el desarrollo de herramientas y soluciones en diversas áreas del conocimiento, sin dejar de lado las problemáticas sociales y económicas que se viven en la actual sociedad.

Es por esto que el presente proyecto se gesta como una iniciativa para poner a disposición de usuarios y microempresas un portal web con un sistema de recomendación de información que permita a los pequeños emprendedores darse a conocer dentro del mercado, y por otra parte brindarle al usuario la posibilidad de conocer un lugar donde posiblemente se suplan sus necesidades.

Partiendo de esta perspectiva se realizó un estudio de importantes referentes teóricos relativos al tema, con el objetivo de conocer el funcionamiento de los sistemas de recomendación para poder desarrollar la mejor propuesta posible; estos referentes de carácter histórico y teórico se presentan a continuación.

1) Sistema de Recomendación de Películas [36]

Modalidad: Artículo.

Institución: Escuela Superior Politécnica del Litoral, Guayaquil - Ecuador

⁸ Técnico especialista de Lotus en el International Technical Support Organization Center en las oficinas centrales de la compañía en Cambridge.

En este proyecto se pretende hacer uso de sistemas de recomendación para seleccionar entre un gran volumen de ítems aquello que un usuario en particular necesita. Para ello y dependiendo de la naturaleza del problema, existen diversos objetivos de los sistemas recomendadores: filtrado, búsqueda y recomendación.

Objetivo: Implementar un Sistema de Recomendación de Películas basado en el paradigma MapReduce para el procesamiento masivo de datos, utilizando el dataset de Netflix y evaluar los resultados de las recomendaciones presentadas por el sistema para determinar su grado de exactitud con respecto al gusto del usuario. En este proyecto se utilizó el método de Filtrado colaborativo basado en ítems. La implementación de los algoritmos se desarrolló en Python. Se usaron los ratings o puntuaciones que los usuarios han dado a las películas.

Análisis: A Partir de las bases de datos de netflix se obtuvo un conjunto de archivos que describen la estructura que describe a los usuarios y a los ítems, en este caso personas y películas.

La descripción de la estructura de los archivos en la base de datos es:

Usuario Id - Película Id - Puntuación

Donde:

- Usuario Id es un dato numérico que identifica a cada usuario dentro del sistema.
- Película Id es un dato que identifica únicamente a cada película.
- Puntuación es el valor que los usuarios le dan a las películas.

2) **Sistema de recomendación de restaurantes georreferenciados [37]**

Modalidad: proyecto de grado

Institución: Universidad de Jaén, España, Septiembre, 2008

Introducción: Este proyecto se ha centrado en los Sistemas de Recomendación y los servicios de Geocodificación de direcciones.

Problema: Cuando un turista llega a su destino puede que tenga mucha información sobre qué lugares visitar, hoteles para hospedarse, y restaurantes para comer. Pero esta información es general, no está basada en sus propios gustos y encuentra sus necesidades un tanto desatendidas ya que normalmente desconoce la mayoría de los servicios que se ofrecen, la situación geográfica de éstos o incluso el cómo llegar hacia ese servicio.

Por tanto, el uso de los sistemas de recomendación-georreferenciados puede permitir al turista visitar aquellos lugares, hospedarse en aquellos hoteles o comer en aquellos restaurantes que mejor se adaptan a sus propios gustos y conocer con un simple clic de ratón la situación geográfica exacta de los restaurantes y de los servicios que les rodean.

El proyecto tiene como fin el desarrollo de un prototipo de software que se denominará REJA (Restaurantes de Jaén) que permitirá georreferenciar y realizar recomendaciones de los restaurantes de la provincia de Jaén.

Objetivo:

- Generar un sistema de recomendación y geolocalización de restaurantes para la provincia de Jaén.
- instalar el software en cualquier Web turística que permita a los usuarios recibir recomendaciones sobre qué restaurantes acudir cuando visiten Jaén, o cualquier pueblo de la provincia atendiendo a sus propios gustos.
- mejorar la experiencia del visitante, de manera que se sienta tan cómodo en el territorio que visita como en el suyo propio.

3) Sistema de recomendación personalizada de contenido vídeo [38]

Objetivo: Con este proyecto se quiere mejorar uno de los productos principales de la empresa, que consiste en un sistema de televisión interactiva que permite a los usuarios poder disfrutar de varios servicios tales como una aplicación ofimática, una guía local y la adquisición de contenidos de distintos tipos: películas, series, clips, música, información cultural y videojuegos entre otros. Dado que no existe ningún servicio de personalización en el sistema y la relevancia que tiene para los usuarios el servicio de adquisición de contenidos de vídeo, se pretende mejorar este servicio añadiéndole personalización. De manera que se quiere desarrollar un sistema de recomendación de contenido de vídeo que funcione de forma complementaria al servicio de adquisición de contenidos. Para ello será necesario diseñar un sistema de recomendación de contenido de vídeo adecuado al entorno del sistema e integrarlo en forma de nuevas funcionalidades del sistema.

4) Prototipo Web de un Sistema de Recomendación Recursos Docentes [39]

Objetivo: Análisis, diseño e implementación de un prototipo de aplicación Web para la recomendación de recursos docentes, a partir de las representaciones internas de alumnos y recursos, así como de las valoraciones realizadas por los estudiantes.

5) Aplicaciones Web comerciales similares:

Qype [40] [41]: <http://www.qype.es/> : Este sistema tiene una base de datos sobre los comentarios, fotos y reseñas creada por los usuarios sobre los establecimientos comerciales. Con una comunidad activa de Qypers revisando y recomendando lugares nuevos cada día, también es una red cada vez más importante para las empresas que se promocionan y gestionan su reputación online. El competidor Yelp anunció la adquisición de Qype3 el 24 de octubre de 2012 y las recomendaciones se fusionaron con la página de Yelp en octubre de 2013.

Yelp [42] [43] [44]: <http://www.yelp.es/> :

Yelp es una guía urbana y electrónica la cual se fundó en el 2004, ayuda a la gente a encontrar los mejores sitios para comer, hacer compras, beber, relajarse y divertirse. Se basa en las opiniones argumentadas de una comunidad y de residentes locales. Yelp tiene aplicación tanto en IOS, Android y BlackberryOS. Yelp se enfocó en el mercado comercial norteamericano pero desde el 2012 al comprar Qype por 50 millones de dólares ya se extiende por Europa y América del sur.

Ciudad Gurú [45]: <http://www.ciudadguru.com.co/> : Gurú es una comunidad de recomendaciones de Colombia. Brinda una guía completa de entretenimiento de una ciudad en concreto. La información que brinda es acerca de restaurantes, de las películas en cartelera y de lugares sofisticados de las principales ciudades del país o encontrar información sobre un lugar específico como una ferretería o droguería.

Otras aplicaciones web comerciales:

- www.facebook.com
- www.twitter.com
- movies.netflix.com
- www.youtube.com
- www.lastfm.es
- www.amazon.com

1.9.3 MARCO CONCEPTUAL

Filtrado colaborativo: Es un conjunto de algoritmos que utilizan las valoraciones de los usuarios sobre ciertos elementos de un conjunto total de elementos para predecir variaciones en el resto de los elementos y recomendar los de mayor valoración. [46]

Establecimiento comercial: Se entiende por Establecimiento de Comercio un conjunto de bienes organizados por el empresario para realizar los fines de la empresa. Una misma persona natural o jurídica podrá tener varios establecimientos de comercio, estos podrán pertenecer a uno o varios propietarios, y destinarse al desarrollo de diversas actividades comerciales. [47]

Microempresa: La microempresa es una planta de personal no superior a los diez trabajadores con activos totales excluida la vivienda por valor inferior a quinientos salarios mínimos mensuales legales vigentes.⁹

TIC: Las TIC o Tecnologías de la Información y Comunicación por sus siglas, son aquellas tecnologías que son usadas para la gestión y transformación de la información mediante el uso de ordenadores y programas que permiten crear, modificar, almacenar, administrar, proteger y recuperar esa información.

Back-end Un programador tiende a ser backend. Es la labor de ingeniería que compone el acceso a bases de datos y generación de plantillas del lado del servidor. En backend se encargan de implementar herramientas como MySQL, Postgres, SQL Server o MongoDB. Luego, un lenguaje como PHP o JSP, o frameworks como RoR, Django, Node.JS o .NET se conectan a la base de datos. A través de estos lenguajes y frameworks se recibe, procesa y envía información al navegador del usuario. En código HTML (que crea el frontend) o enviando datos puros en XML, RSS o JSON, para ser procesados por Javascript. [48]

Front-end: De forma general, front-end hace referencia al estado inicial de un proceso. Contrastá con back-end, que se refiere al estado final de un proceso. La idea general es que el front-end es responsable de recoger entradas de los usuarios, y ser procesadas de tal manera que cumplan las especificaciones para que el back-end pueda usarlas. La conexión entre front-end y el back-end es un tipo de interfaz. [49]

Organización: Una organización es un sistema comercial diseñado y creado para lograr metas mediante la gestión del talento humano. Están compuestas por subsistemas interrelacionados que cumplen diferentes funciones con el fin de cumplir sus objetivos.

⁹ Ley 905 de 2005, “promoción del desarrollo de la micro, pequeña y mediana empresa colombiana”

Framework: La palabra inglesa "framework" (marco de trabajo) define, en términos generales, un conjunto estandarizado de conceptos, prácticas y criterios para enfocar un tipo de problemática particular que sirve como referencia, para enfrentar y resolver nuevos problemas de índole similar. Son diseñados con la intención de facilitar el desarrollo de *software*, permitiendo a los diseñadores y programadores pasar más tiempo identificando requerimientos de *software* que tratando con los tediosos detalles de bajo nivel de proveer un sistema funcional.

Tecnologías web: Las tecnologías web son un conjunto de herramientas que facilitan el desarrollo y la administración de un sitio web.

1.9.4 MARCO METODOLÓGICO

La metodología seleccionada para el desarrollo del prototipo del sistema de recomendación corresponde a la metodología UP, lo que define un conjunto de pasos o actividades a seguir que permiten el desarrollo eficaz de diferentes tipos de sistemas de software, para diferentes áreas de aplicación, diferentes tipos de organizaciones, diferentes niveles de competencia y diferentes tamaños de proyectos, lo que provee una potencial ventaja de su uso en este proyecto. Adicionalmente la metodología se encuentran divididas en las fases de todo proyecto de ingeniería (Requerimientos, análisis, diseño implementación y pruebas.)

En la siguiente tabla se muestra la aplicación de la metodología al desarrollo del proyecto.

ETAPAS DEL SISTEMA PROYECTO		METODOLOGÍA	ACTIVIDADES	DÍAS
Modelo del negocio	Sistemas de recomendación	U.P	Modelo de Dominio Glosario de términos	
	Portal Web		Modelo de Dominio Glosario de términos	
	Sistema de recomendación	U.P	Identificación de actores Lista preliminar de casos de uso Depuración de casos de uso Documentación de casos de uso	30
	Portal Web		Identificación de actores Lista preliminar de casos de uso Depuración de casos de uso	

			Documentación de casos de uso	
Análisis	Sistema de recomendación	U.P	Diagramas de secuencia	66
			Diagramas de colaboración	
			Diagramas de actividad	
			Diagramas de estado	
			Diagramas de clase	
	Portal Web	U.P	Diagramas de secuencia	66
			Diagramas de colaboración	
			Diagramas de actividad	
			Diagramas de estado	
			Diagramas de clase	
Diseño	Sistema de recomendación	U.P	Lista inicial de clases	68
			Responsabilidades de las clases	
			Modelo de Interfaz	
			Modelo Lógico	
			Modelo Físico	
	Portal Web	U.P	Modelo Relacional	68
			Diccionario de Datos	
			Lista inicial de clases	
			Responsabilidades de las clases	
			Modelo de Interfaz	
Implementación	Sistema de recomendación	U.P	Modelo Lógico	73
			Modelo Físico	
			Modelo Relacional	
	Portal Web	U.P	Diccionario de Datos	73
			Definir la organización del código	
			Implementar Clases	
Pruebas	Sistema de recomendación	U.P	Documentación del código Fuente	8
			Definir la organización del código	
	Portal Web	U.P	Implementar Clases	
			Documentación del código Fuente	
			Prueba de ejecución	
Despliegue	Portal Web	U.P	Pruebas de Error medio absoluto MAE	8
			Prueba de Sistema	
			Prueba de Integración	
			Diagramas de componentes Diagramas de paquetes Diagramas de despliegue	

Tabla 6. Aplicación de la metodología

1.10 FACTIBILIDAD

1.10.1 FACTIBILIDAD TÉCNICA

Con base en la información obtenida, la experiencia técnica con sistemas de información, aplicaciones web y bases de datos, y las herramientas tecnológicas seleccionadas se puede decir que el desarrollo del presente proyecto es técnicamente posible.

1.10.2 FACTIBILIDAD ECONÓMICA

A continuación se presenta una tabla que resume la factibilidad económica.

ITEM	CANTIDAD	VALOR UNITARIO	SUBTOTAL
Recursos tecnológicos			
Computador de escritorio.	2	\$ 1'600.000	\$ 3'200.00
Servidor	1	\$ 2'500.000	\$ 2'500.000
Impresora	1	\$ 350.00	\$ 350.000
Sistema operativo Linux	1	\$ 0	\$ 0
SGBD PosgreSQL	1	\$ 0	\$ 0
Django	1	\$ 0	\$ 0
Bootstrap	1	\$ 0	\$ 0
Servidor Web Nginx	1	\$ 0	\$ 0
Recursos materiales y otros			
Papelería	1	\$ 7.000	\$ 100.000
Internet		\$ 80000 Mensual	\$ 720.000
Recurso humano		Costo por día	
Desarrollador de software	1	\$ 30.000	\$ 7'350.000
Consultor	2	\$ 30.000	\$ 14'700.000
		COSTO TOTAL	\$ 18'135.000

Tabla 7. Factibilidad económica

1.10.3 FACTIBILIDAD LEGAL

El software que se utiliza para la implementación del prototipo corresponde a herramientas libres, y los equipos para el desarrollo del proyecto son recursos personales y recursos del grupo de investigación METIS de la Universidad Distrital.

En la siguiente tabla se resumen las licencias correspondientes a las herramientas a utilizar en este proyecto.

Software	Licencia
JQuery, Ubuntu server Linux.	GPL
Twitter Bootstrap	Apache 2.0
HTML, CSS.	Creative commons
Python	Python Software Foundation License
PostgreSQL, Django, Redis, Celery	Licencia BSD
Sqlite	Dominio publico
Spatialite	MPL GPL LGPL
PostGIS	GNU
Git	GNU General Public License v2, GNU Lesser General Public License 2.1
RabbitMQ	Mozilla Public License
Nginx	Segunda cláusula de la licencia BSD

Tabla 8. Factibilidad legal

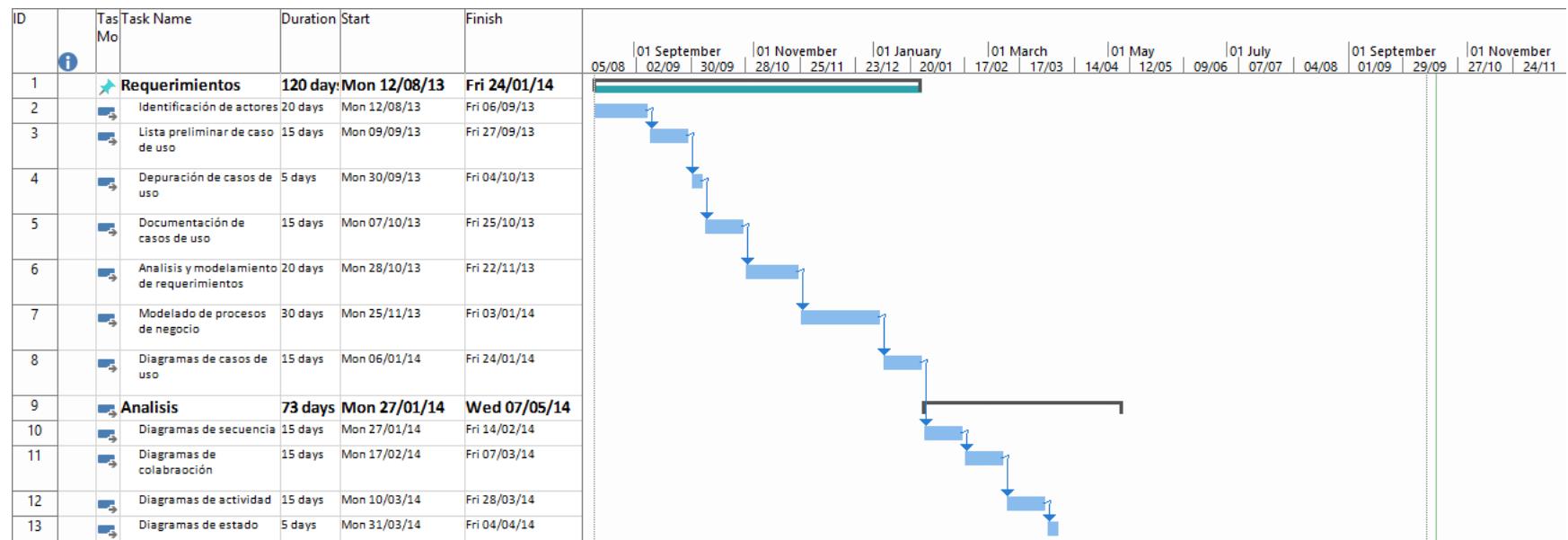
1.10.4 FACTIBILIDAD OPERATIVA

Para el desarrollo del módulo de búsqueda se hace necesaria la disponibilidad de los siguientes recursos humanos:

- Desarrolladores de software con conocimiento en lenguajes de programación orientados a aplicaciones web y conocimiento en el manejo de motores de bases de datos.

De acuerdo con lo anterior el recurso humano requerido se encuentra disponible y capacitado con los conocimientos necesarios para desarrollar el proyecto. Por otro lado, los recursos técnicos necesarios para la implementación del proyecto se encuentran disponibles por lo que el proyecto es factible a nivel operativo.

1.11 CRONOGRAMA



2 MODELO DEL NEGOCIO

2.1 MODELO DE PROCESOS

A continuación se presenta la funcionalidad general de casa sistema, se escogieron los principales diagramas, todos los demás los encuentra en el anexo Modelamiento del sistema,

2.1.1 *Modelo de procesos Sistema de establecimientos*

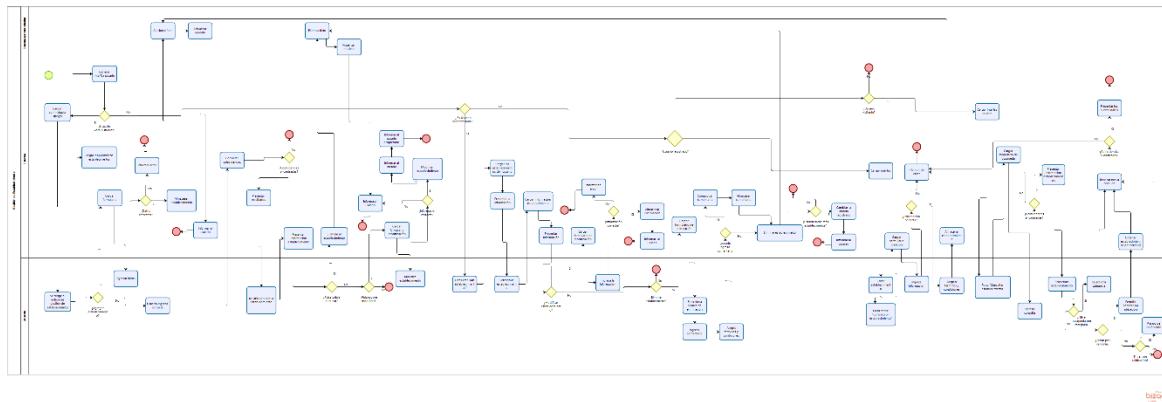


Ilustración 6. *Modelo de procesos Gestión de establecimientos*

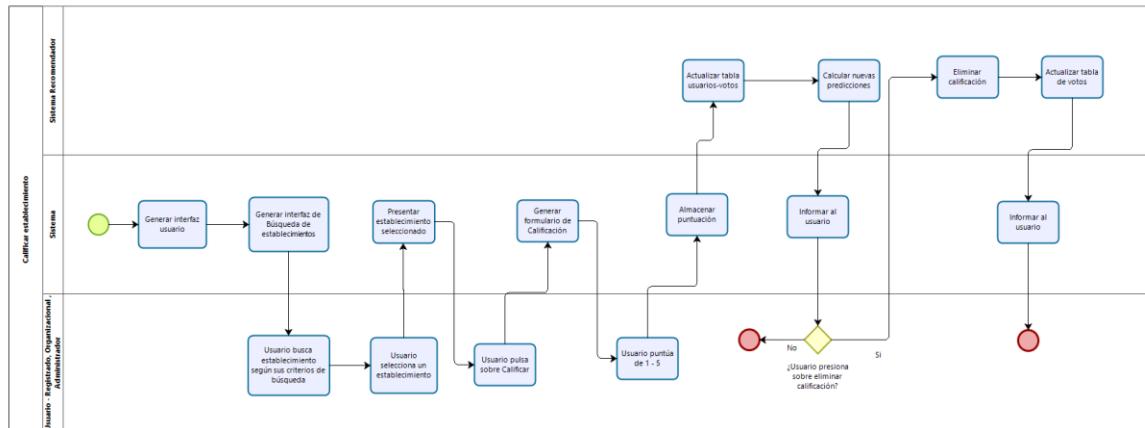
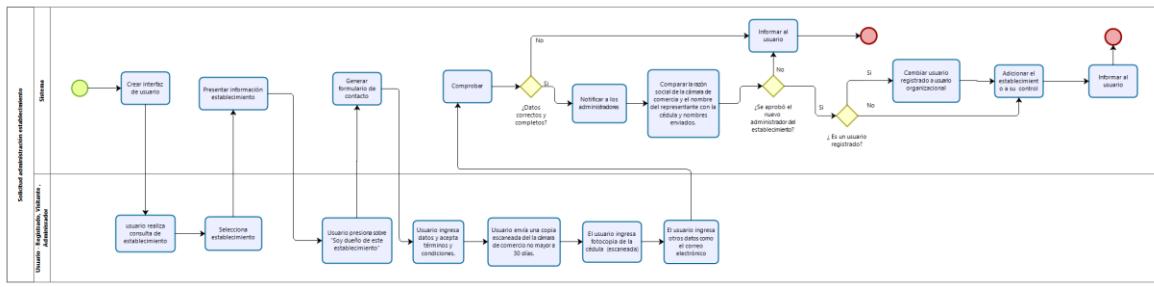
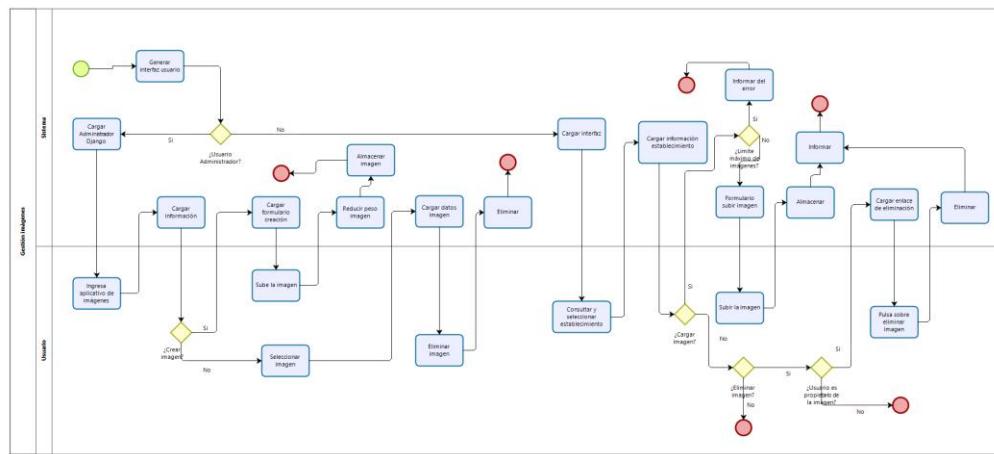


Ilustración 7. *Modelo de procesos Calificar establecimiento*



Powered by
bizagi
Process Automation

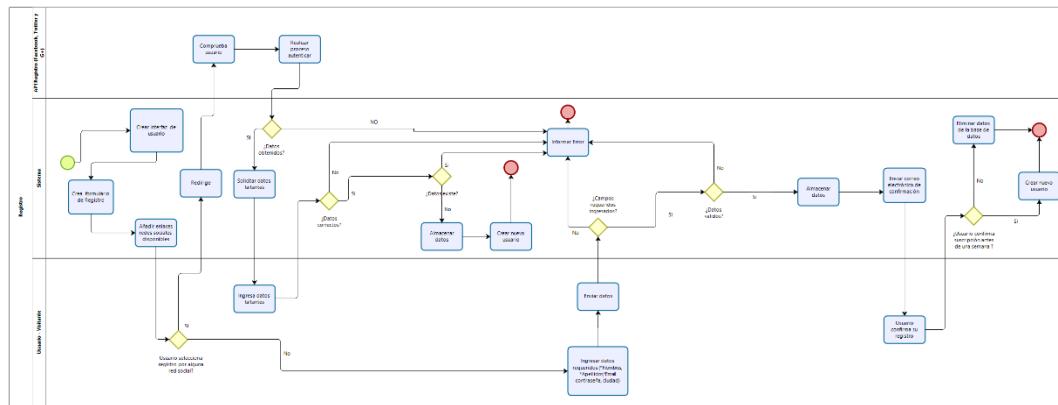
Ilustración 8. Modelo de procesos Solicitud administración



Powered by
bizagi
Process Automation

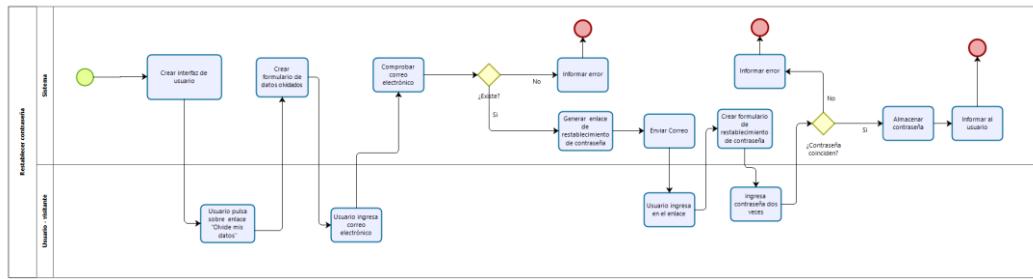
Ilustración 9. Modelo de procesos Gestión de imágenes

2.1.2 Modelo de procesos Sistema de usuarios



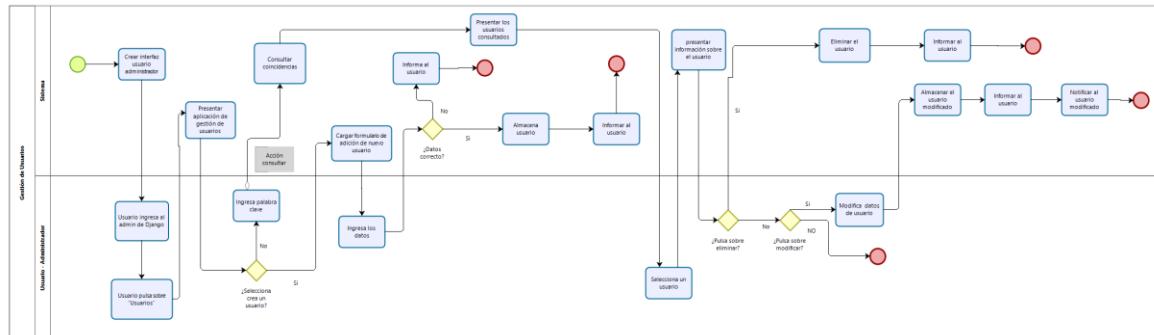
bizagi
Business Process Management

Ilustración 10. Modelo de procesos Registro



bizagi
Business Process Management

Ilustración 11. Modelo de procesos Restablecer contraseña



bizagi
Business Process Management

Ilustración 12. Modelo de procesos Gestión de usuarios

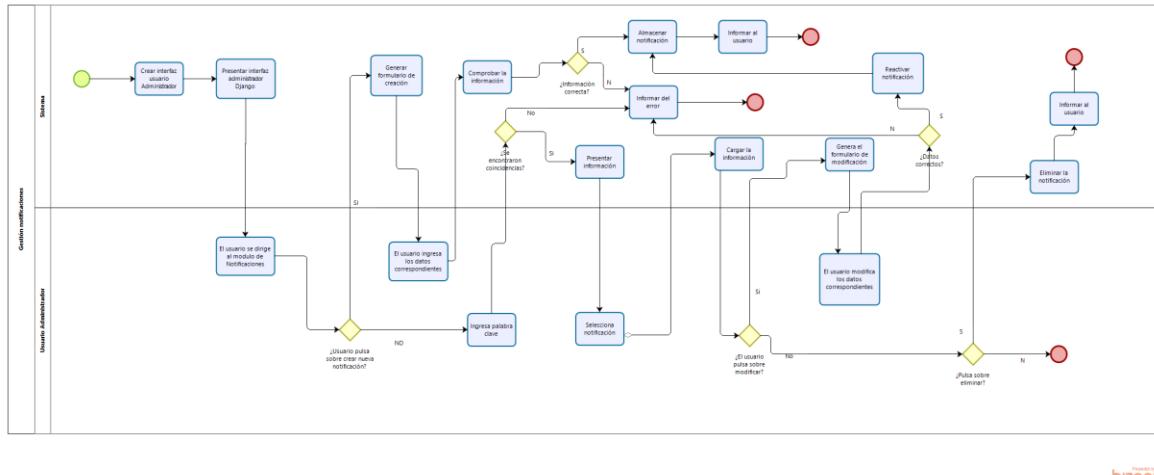


Ilustración 13. Modelo de procesos Gestión de notificaciones

2.1.3 Modelo de procesos sistema de Recomendación

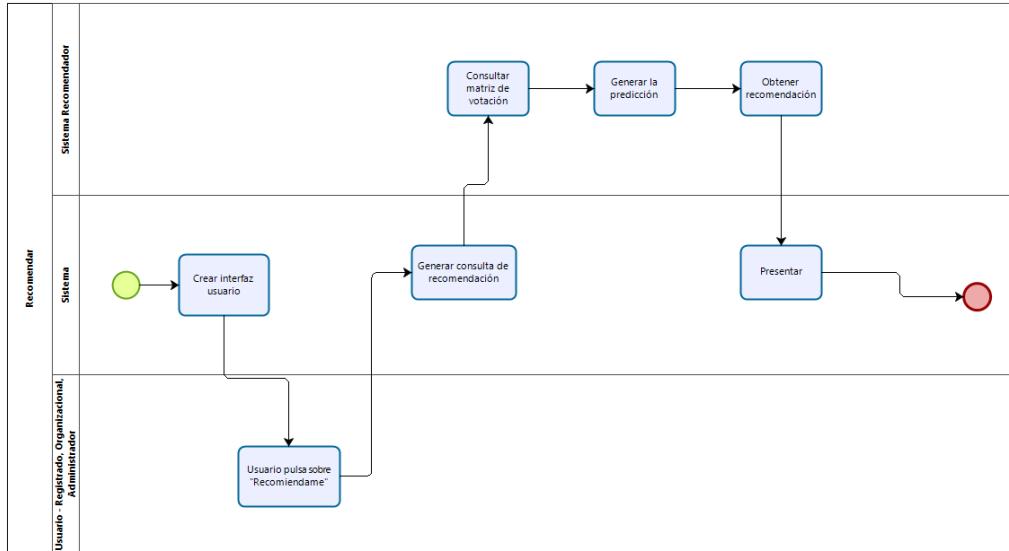


Ilustración 14. Modelo de procesos Recomendar

2.2 MODELOS DE DOMINIO

A continuación se presenta el Modelo de Dominio para cada sistema.

2.2.1 Modelo de Dominio sistema de establecimientos

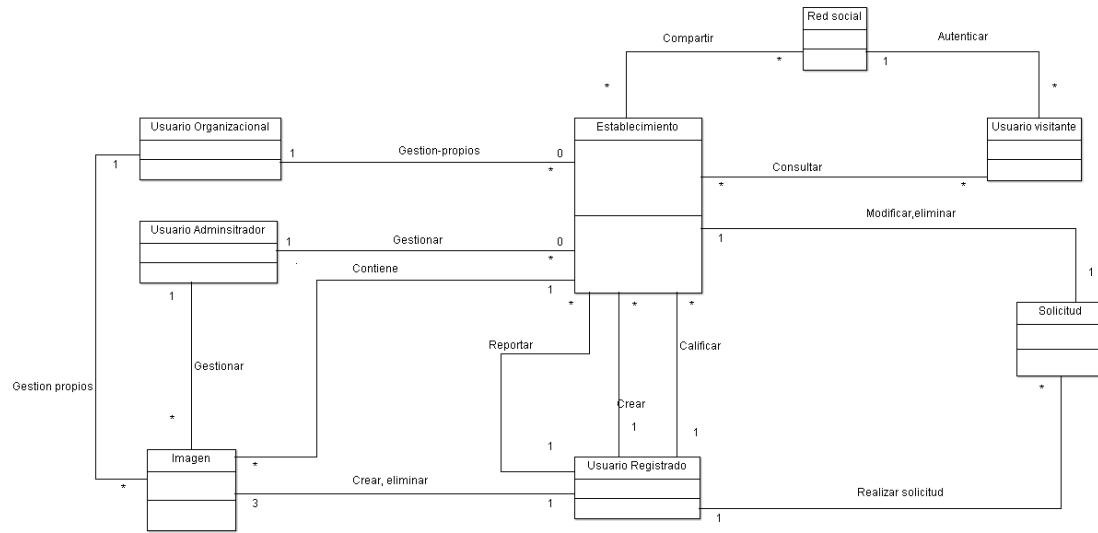


Ilustración 15. Modelo de dominio Sistema de establecimientos

2.2.2 Modelo de Dominio sistema de usuarios

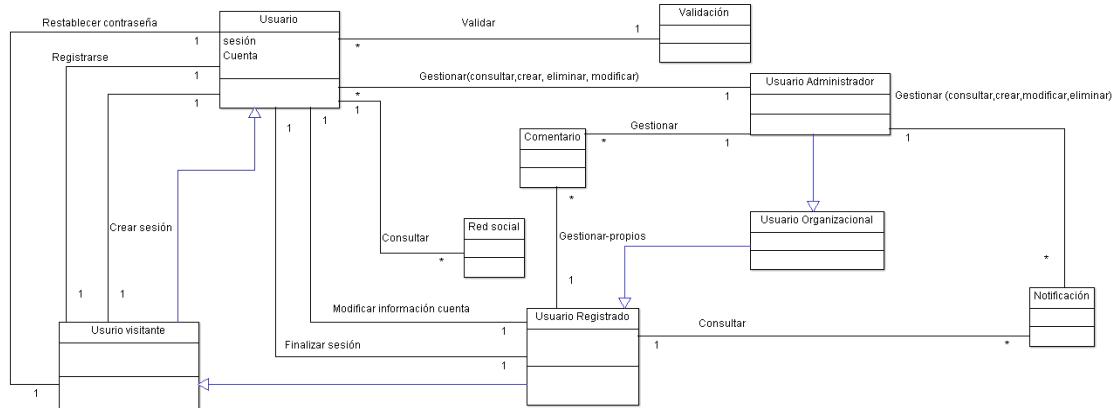


Ilustración 16. Modelo de dominio Sistema de usuarios

2.2.3 Modelo de Dominio sistema Recomendador

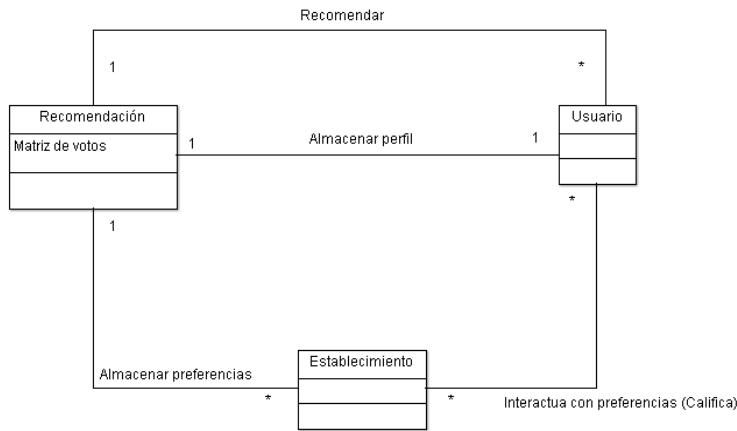


Ilustración 17. Modelo de dominio Sistema recomendador

2.3 GLOSARIO DE TÉRMINOS.

Concepto	Descripción
Recomendación	Es el sistema de Recomendación, se encarga de modelar perfiles de diferentes tipos de usuarios para posteriormente obtener una predicción de selección sobre algún ítem y generar por último la recomendación.
Recomendar	Es la acción del Sistema Recomendador que a partir de su modelo asiste al usuario en la selección de algún ítem.
Usuario	Es el usuario genérico del sistema, los demás tipos de usuarios heredan sus atributos y métodos.
Sesión	Es el cuadro espacio-tiempo en el que un usuario accede y utiliza los servicios del portal.
Usuario Administrador	Es el actor encargado de controlar y gestionar usuarios y establecimientos, además tiene las mismas capacidades de acceso que los demás actores.
Usuario Invitado	Es el usuario no identificado que inicialmente entra en el portal.
Usuario Registrado	Es el actor que puede acceder a las herramientas de recomendación del sistema.
Usuario Organizacional	Es el actor que administra establecimientos de su propiedad, además tiene las mismas capacidades de acceso que el actor usuario registrado.
Establecimiento	Es la abstracción de un espacio físico donde se ofrecen bienes económicos (servicios o mercancías) para su venta al público. Lo que se considera como ítem en el Sistema de Recomendación es un establecimiento.
Imagen	Son las imágenes pertenecientes a cada establecimiento, las cuales son creadas por los mismos usuarios.
Red Social	Se encarga de la autenticación por medio del uso de redes sociales (Twitter, Facebook, G+).
Comentario	Es una descripción u opinión textual realizada por el usuario con respecto a un establecimiento.
Validación	Es un formulario encargado de verificar y validar toda la información creada e ingresada por los usuarios del sistema.

Solicitud	Es una petición realizada por los usuarios para Administrar, eliminar o modificar un establecimiento según corresponda, esta será aprobada o desaprobada por un usuario Administrador y el usuario será informado.
Notificación	Son todos los mensajes y alertas que se generan en la plataforma para informar a todos los usuarios. Estos son creados por el usuario Administrador.
Interfaz	Es la forma en que los usuarios pueden comunicarse con el portal, y comprende todos los puntos de contacto entre el usuario y el sistema.
Contraseña	Es una forma de autenticación que utiliza información secreta para controlar el acceso hacia algún recurso informático en este caso el acceso al portal.
Categoría Establecimiento	Es la categoría al que pertenece un establecimiento, los establecimientos comerciales pueden clasificarse según diversos criterios, siendo los más comunes los referidos a su carácter individual o colectivo, a su tamaño, al tipo de productos que comercializa y, principalmente, al sistema de ventas.
Nombre del usuario	Es el nombre de usuario (nombre, apellido) de quien accede a usar los servicios del portal.
E-mail	Es el correo electrónico del usuario que accede a usar los servicios de la plataforma, de la misma manera representa el nombre de ingreso del usuario al portal.
Certificado cámara de comercio	El certificado de Cámara y Comercio, mejor conocido como Certificado de Existencia y Representación Legal, informa a los usuarios si una empresa está activa, quien es el Representante Legal, Gerente, quiénes conforman la junta directiva, cuál es su objeto social, qué duración tiene la empresa, domicilio, fecha de creación, si se han realizado cambios a la estructura de la empresa, etc.
Matriz de Votos	Es una matriz bidimensional que contiene números reales entre 0-5, compuesta del número de usuarios por el número de ítems, representa el modelo del Sistema de Recomendación.
Formulario	Es el conjunto de datos llenados por el usuario para ser enviados y luego ser procesados por el sistema para devolver un resultado al usuario.

Tabla 9. Glosario de términos

3 FASE DE REQUERIMIENTOS

3.1 REQUERIMIENTOS FUNCIONALES

- El portal web debe validar automáticamente la información posterior a su envío. Este proceso, se deben tener en cuenta aspectos tales como obligatoriedad de campos, longitud de caracteres permitida por campo, etc.
- El portal web debe permitir el registro de usuarios, ya sea de manera convencional o mediante el uso de redes sociales tales como Facebook, Twitter o Google+.
- El portal web debe diferenciar entre los diferentes tipos de usuarios, ya sea si es usuario Administrador, Organizacional o Registrado.
- Todos los usuarios deben tener acceso a las notificaciones enviadas por los administradores.
- El Portal debe permitir a los usuarios Administradores gestión de usuarios, comentarios, imágenes, establecimientos, solicitudes, categorías, sub-categorías y notificaciones.
- El portal debe permitir a los usuarios Organizacionales la gestión de establecimientos que le pertenezcan.
- El portal debe permitir a los usuarios registrados crear establecimientos, solicitudes y comentarios, así como gestionar imágenes propias.
- El portal debe permitir a cualquier tipo de usuario expresar sus gustos e intereses entre un número de 1 a 5, donde uno representa poco conforme y 5 muy conforme.
- El portal debe permitir la consulta de establecimientos median su ubicación.
- El portal debe representar la ubicación de los establecimientos en el mapa.
- A partir las puntuaciones otorgados por los usuarios el Sistema de recomendación deberá recomendar un máximo de 10 establecimientos.

3.2 REQUERIMIENTOS NO FUNCIONALES

- Navegador Web como Google Chrome en su versión 37.0.2062.124 o superior, o Firefox en su versión 32.02 o superior.
- Entorno de ejecución Python 2.7
- Pip y easy_install como sistema de gestión de paquetes para Python.
- Django-framework en su versión 1.6.6
- El puerto TCP 80 y 8000 para el servidor de Django y NGINX.
- El puerto TCP 5432 para Postgresql
- El puerto 6379 para Redis.
- Sistema operativo de Ubuntu server 14.04.
- Repositorios de Ubuntu server 14.04

- Servicio de internet con una velocidad de transmisión mínima de 1 Mbps y sin bloqueos por proxy hacia los servidores de los repositorios.

3.3 DEFINICIÓN DE ACTORES DEL SISTEMA

Actor	Descripción
Usuario visitante	Es el actor que inicialmente llega al portal, tiene acceso nulo del sistema de Recomendación, pero puede consultar establecimientos y compartirlos, además puede registrarse para escalar al actor “Usuario Registrado”.
Usuario Registrado	Es un usuario visitante que se registró en el sistema, es decir que tiene un usuario y una clave de acceso para acceder al portal. Puede acceder a las herramientas ofrecidas por el sistema de Recomendación para obtener una recomendación basados en el modelo de perfil, generado a partir de sus calificaciones otorgadas a los establecimientos del sistema, puede realizar solicitudes de creación, modificación y eliminación de establecimientos comerciales, así como también solicitudes de administración.
Usuario Organizacional	Además de las mismas capacidades de acceso del actor “usuario registrado”, este actor puede administrar los establecimientos comerciales de su propiedad en los que realizó solicitud de administración.
Usuario Administrador	Además de las mismas capacidades de acceso del actor “Usuario Organizacional” este actor está encargado de la gestión de todos los componentes del sistema, desde hacer creaciones, pasando por modificaciones y consultas, hasta llegar eliminaciones de los elementos ofrecidos por el portal.

Tabla 10. Definición de actores del sistema

3.4 LISTA PRELIMINAR DE CASOS DE USO

A continuación se presenta la lista preliminar de casos de uso para cada sistema.

3.4.1 Sistema de usuarios

Actor	Caso de uso
Usuario Visitante	<ul style="list-style-type: none"> • Registro convencional • Registro social • Iniciar sesión • Restablecer contraseña
Usuario registrado	<ul style="list-style-type: none"> • Cerrar sesión • Modificar cuenta • Consultar notificaciones • Gestionar comentarios: <ul style="list-style-type: none"> ○ Crear, eliminar, consultar • Consultar perfil
Usuario organizacional	
Usuario Administrador	<ul style="list-style-type: none"> • Gestión notificaciones <ul style="list-style-type: none"> ○ Crear, modificar, eliminar, consultar • Gestión usuarios <ul style="list-style-type: none"> ○ Crear, modificar, eliminar, suspender, consultar • Gestión comentarios <ul style="list-style-type: none"> ○ Crear, modificar, eliminar, consultar

Tabla 11. Lista preliminar Casos de Usos Sistema de usuarios

3.4.2 Sistema de Establecimientos

Actor	Caso de uso
Usuario Visitante	<ul style="list-style-type: none"> • Consultar establecimiento <ul style="list-style-type: none"> ◦ Seleccionar establecimiento ◦ Filtrar búsqueda por categoría ◦ Filtrar búsqueda por sub-categoría. ◦ Filtrar búsqueda por localización. • Compartir establecimiento
Usuario registrado	<ul style="list-style-type: none"> • Solicitar administración establecimiento. • Solicitar modificación establecimiento. • Solicitar eliminación inexistencia establecimiento. • Solicitar eliminación duplicidad establecimiento. • Gestión imágenes: <ul style="list-style-type: none"> ◦ Crear ◦ Eliminar • Gestión establecimiento. <ul style="list-style-type: none"> ◦ Crear
Usuario organizacional	<ul style="list-style-type: none"> • Gestión establecimiento <ul style="list-style-type: none"> ◦ modificar, eliminar, consultar • Gestión imágenes <ul style="list-style-type: none"> ◦ Crear, eliminar, consultar
Usuario Administrador	<ul style="list-style-type: none"> • Gestión establecimiento <ul style="list-style-type: none"> ◦ Consultar, crear, modificar, eliminar. • Gestión imágenes <ul style="list-style-type: none"> ◦ Consultar, crear, eliminar.

Tabla 12. Lista preliminar Casos de Usos Sistema de establecimientos

3.4.3 Sistema recomendador

Actor	Caso de uso
Usuario Visitante	
Usuario registrado	<ul style="list-style-type: none"> • Calificar • Obtener recomendación • Eliminar calificación
Usuario organizacional	
Usuario Administrador	
Sistema establecimientos	<ul style="list-style-type: none"> • Computar

Tabla 13. Lista preliminar Casos de Usos Sistema recomendador

3.5 MODELOS DE CASOS DE USO

A continuación se presentan los Modelos de Casos de uso para cada sistema.

3.5.1 Modelos de Casos de Uso del sistema de establecimientos

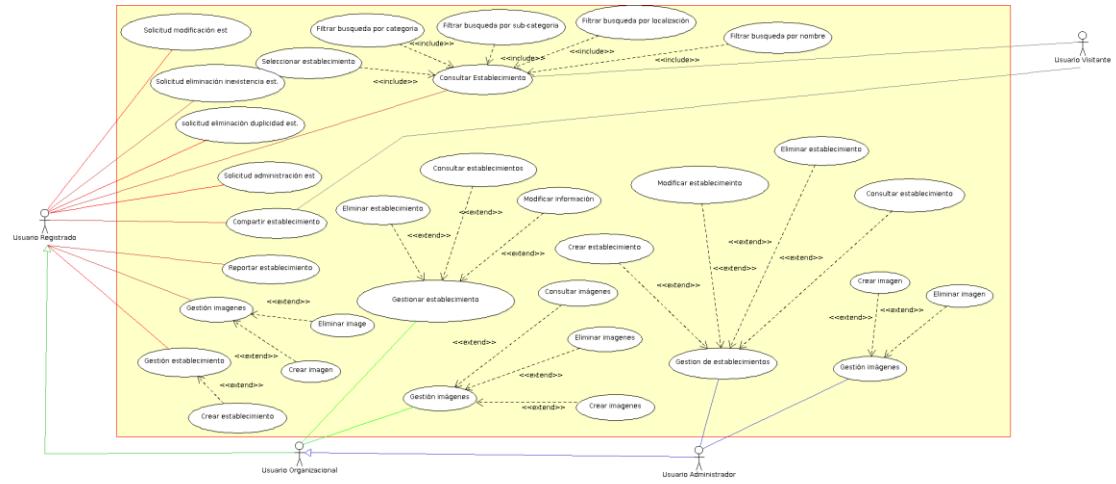


Ilustración 18. Modelo de casos de uso Sistema de establecimientos

3.5.2 Modelos de Casos de Uso del sistema de usuarios

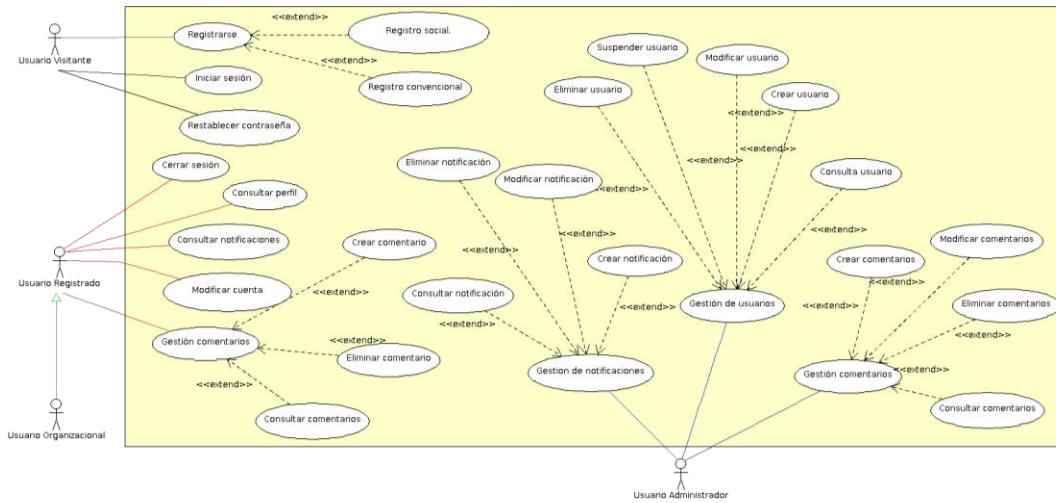


Ilustración 19. Modelo de casos de uso Sistema de usuarios

3.5.3 Modelos de Casos de Uso del sistema Recomendador

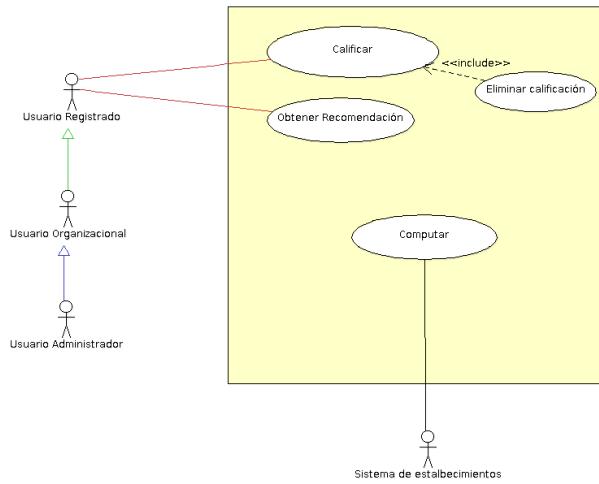


Ilustración 20. Modelo de casos de uso Sistema recomendador

3.5.4 Modelos de Casos de Uso integrado

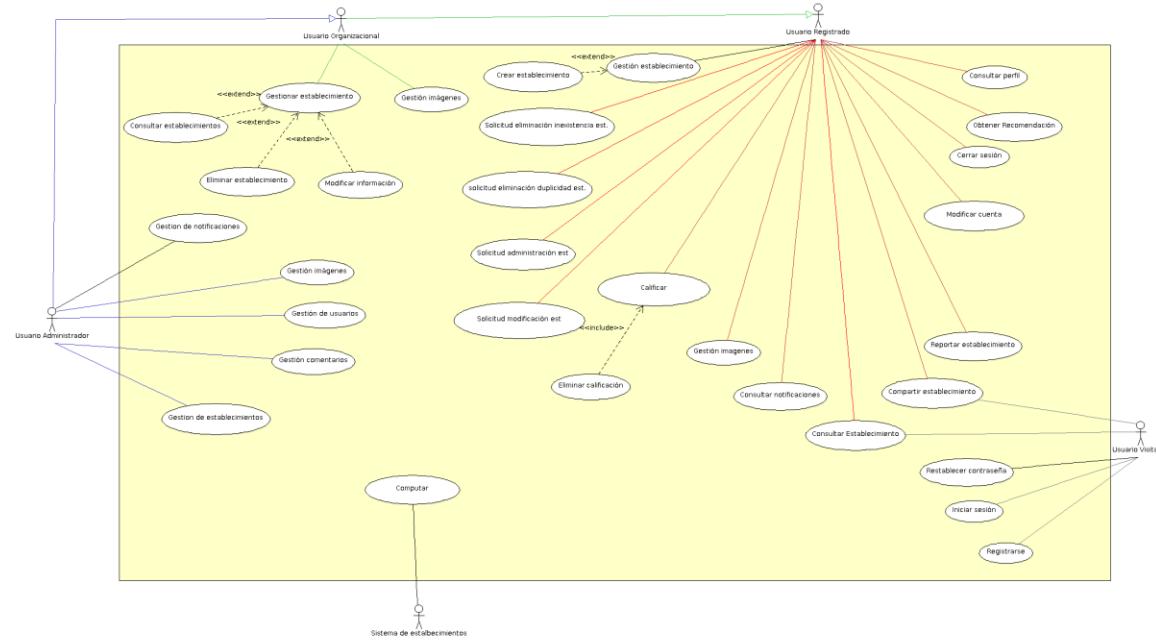


Ilustración 21. Modelo de casos de uso Integrado

3.5.5 Documentación Casos de Uso

A continuación presentamos 3 casos de uso, los demás los encontrará en el anexo de Modelamiento del sistema.

IDENTIFICACION	CASO DE USO	ACTORES
----------------	-------------	---------

01	Registro convencional	Usuario Visitante
OBJETIVO		
Crear un usuario Registrado que pueda hacer uso de los servicios del portal.		
Pre-condiciones	El usuario debe haber ingresado a la página principal.	
Post-condiciones	Se creará un nuevo usuario el cual contará con un nombre de usuario y una contraseña para utilizar los servicios del portal.	
CURSO NORMAL DE LOS EVENTOS		
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA	
1. El usuario pulsa sobre el botón “Registrarse”	2. El sistema carga el formulario Registro.	
3. El usuario ingresa los datos solicitados en el formulario.		
4. Pulsa en “Regístrate”.	5. El sistema comprueba que el nombre de usuario no exista y procede a almacenarlo.	
	6. El sistema genera un Link de confirmación	
	7. Se le envía al correo electrónico del usuario el Link de confirmación.	
	8. Se informa al usuario.	
9. El usuario accede al link de confirmación.	10. El sistema comprueba el link y confirma su registro.	
	11. Se le informa al usuario sobre el éxito de registro.	
	12. Se redirige al usuario a la página principal.	
Flujo alternativo		
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA	
3. El usuario no ingresa algún campo requerido.	5. No se envían los datos hasta que el usuario ingrese los datos requeridos en los campos solicitados. 6. Se envía el mensaje “Error Debe ingresar los campos obligatorios”.	
3. El usuario ingresa datos erróneos.	5. Al comprobarse los datos son incorrectos, se le envía mensaje al usuario sobre el tipo de error generado.	
3. El usuario ingresa un nombre de usuario existente	5. El sistema comprueba que el nombre de usuario existe y envía al usuario el error.	
9. El usuario no confirma su suscripción mediante el link enviado al correo en menos de 30 días	7. El sistema elimina al usuario.	

Puntos de interrupción	
EL usuario puede cancelar en el punto 3	

Tabla 14. Documentación Caso de uso Registro convencional

IDENTIFICACION	CASO DE USO	ACTORES
27	Filtrar búsqueda por localización	Usuario Registrado. Usuario Organizacional. Usuario Administrador Usuario visitante.
OBJETIVO		
Filtrar una consulta realizada a partir de la ubicación geográfica del usuario.		
Pre-condiciones	El usuario debe estar registrado. El usuario debe estar autenticado. El usuario debe realizar una consulta de establecimientos	
Post-condiciones	Se filtran los establecimientos.	
CURSO NORMAL DE LOS EVENTOS		
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA	
1. El usuario mueve su ubicación.	2. El sistema obtiene la información de la ubicación.	
	3. El sistema realiza la consulta de los establecimientos que se encuentra en esa ubicación.	
	5. El sistema presenta la información.	
Flujo alternativo		
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA	
2. No se cargó correctamente el mapa	4. El sistema le informa del error.	
2. No hay establecimiento en esa posición	3. Se informa al usuario	
Puntos de interrupción		

--

Tabla 15. Documentación caso de uso *Filtrado por localización*

IDENTIFICACION	CASO DE USO	ACTORES		
52	Calificar establecimiento	Usuario Registrado. Usuario Organizacional. Usuario Administrador.		
OBJETIVO				
Calificar los establecimientos comerciales mediante el uso de puntuación de 1-5.				
Pre-condiciones	El usuario debe estar registrado. El usuario debe estar autenticado El usuario debe haber seleccionado un establecimiento comercial.			
Post-condiciones	Se calificó un establecimiento comercial. Se actualiza el modelo del perfil de usuario.			
CURSO NORMAL DE LOS EVENTOS				
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA			
1. El usuario pulsa el botón calificar.	2. El sistema carga el formulario de calificación del establecimiento seleccionado.			
3. El usuario califica de uno a cinco, mediante el uso de estrellas.	4. El sistema almacena y cuantifica las puntuaciones proporcionadas por el usuario.			
	5. Actualiza el perfil del usuario en la matriz de votaciones.			
Flujo alternativo				
ACCION DEL ACTOR	RESPUESTA DEL SISTEMA			
Puntos de interrupción				
Se puede cancelar la consulta en los puntos 1				

Tabla 16. Documentación del caso de uso *Calificar establecimiento*

4 FASE DE ANÁLISIS

4.1 DIAGRAMAS DE SECUENCIA

A continuación se presentan los diagramas de secuencia para cada sistema, todos los diagramas los encontrará en el anexo virtual Modelamiento del sistema.

4.1.1 Sistema de usuarios

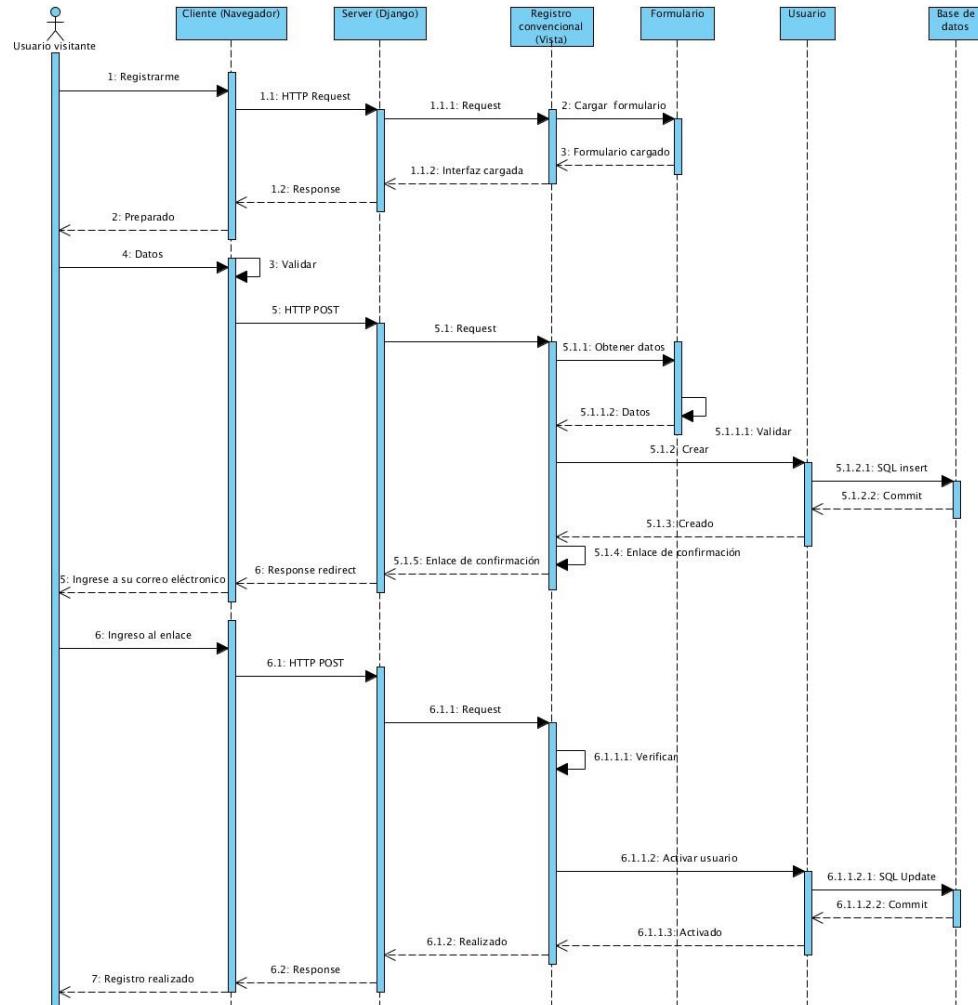


Ilustración 22. Diagrama secuencia Registro convencional

4.1.2 Sistemas de establecimientos

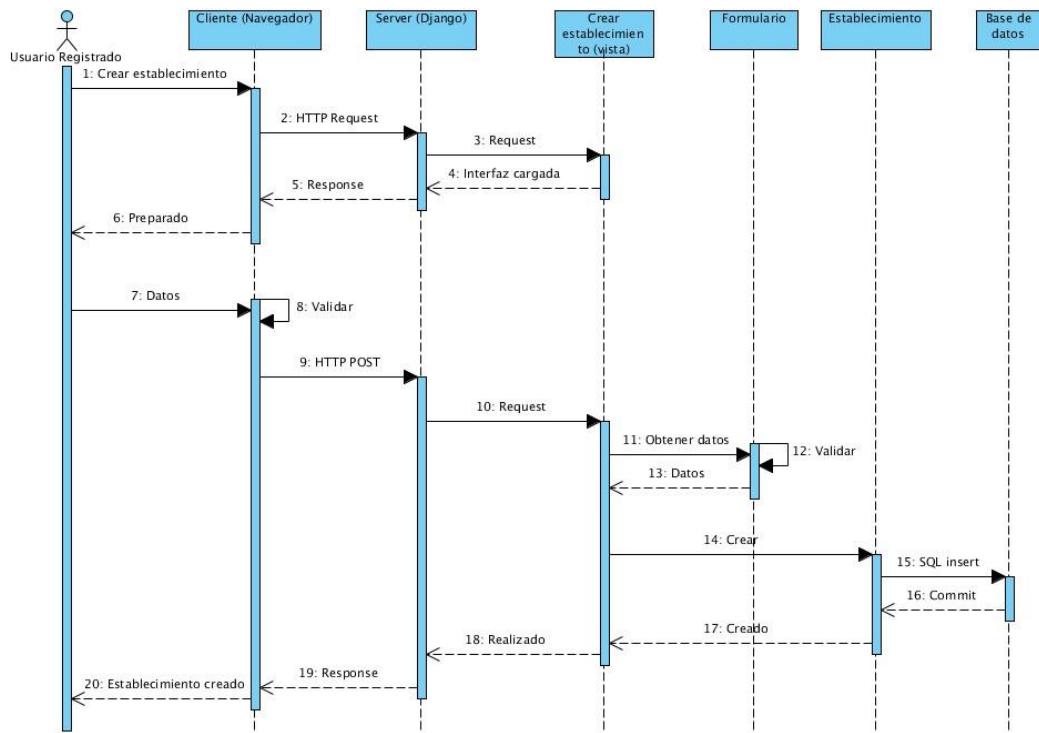


Ilustración 23. Diagrama secuencia Crear establecimiento

4.1.3 Sistema recomendador

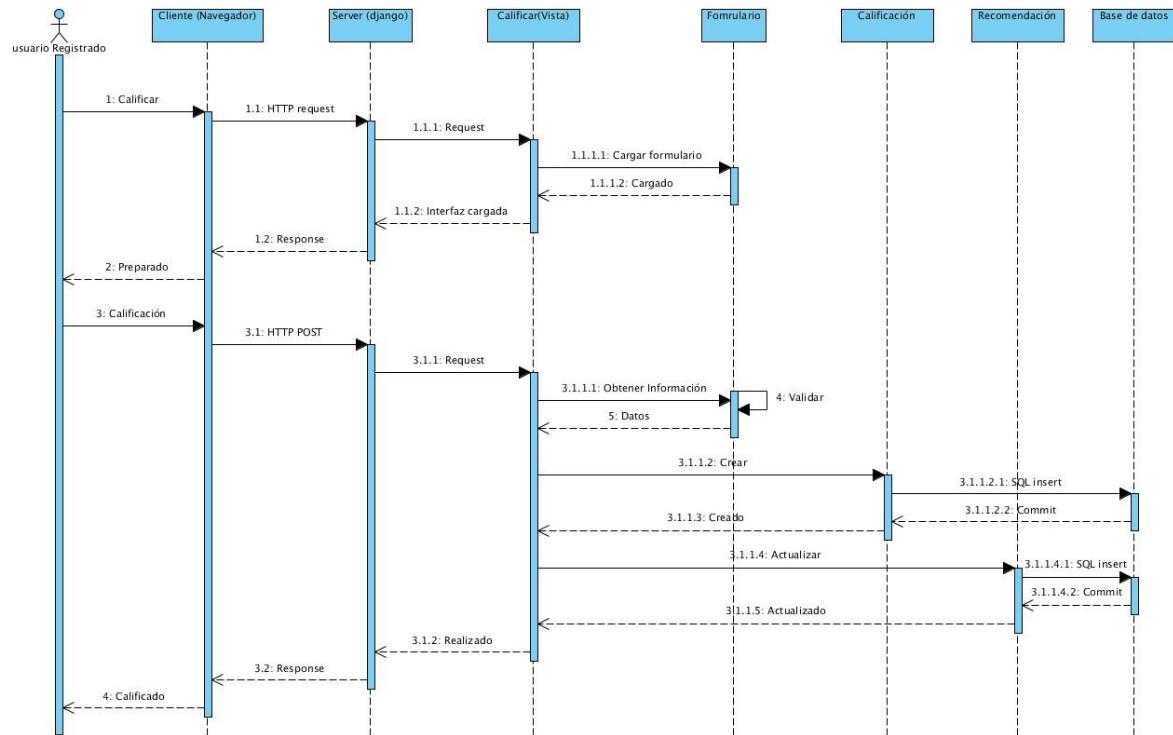


Ilustración 24. Diagrama de secuencia Calificar establecimiento

4.2 DIAGRAMA DE ACTIVIDAD

A continuación se presentan los diagramas de actividad para cada sistema, todos los diagramas los encontrará en el anexo virtual Modelamiento del sistema.

4.2.1 Sistema de usuarios

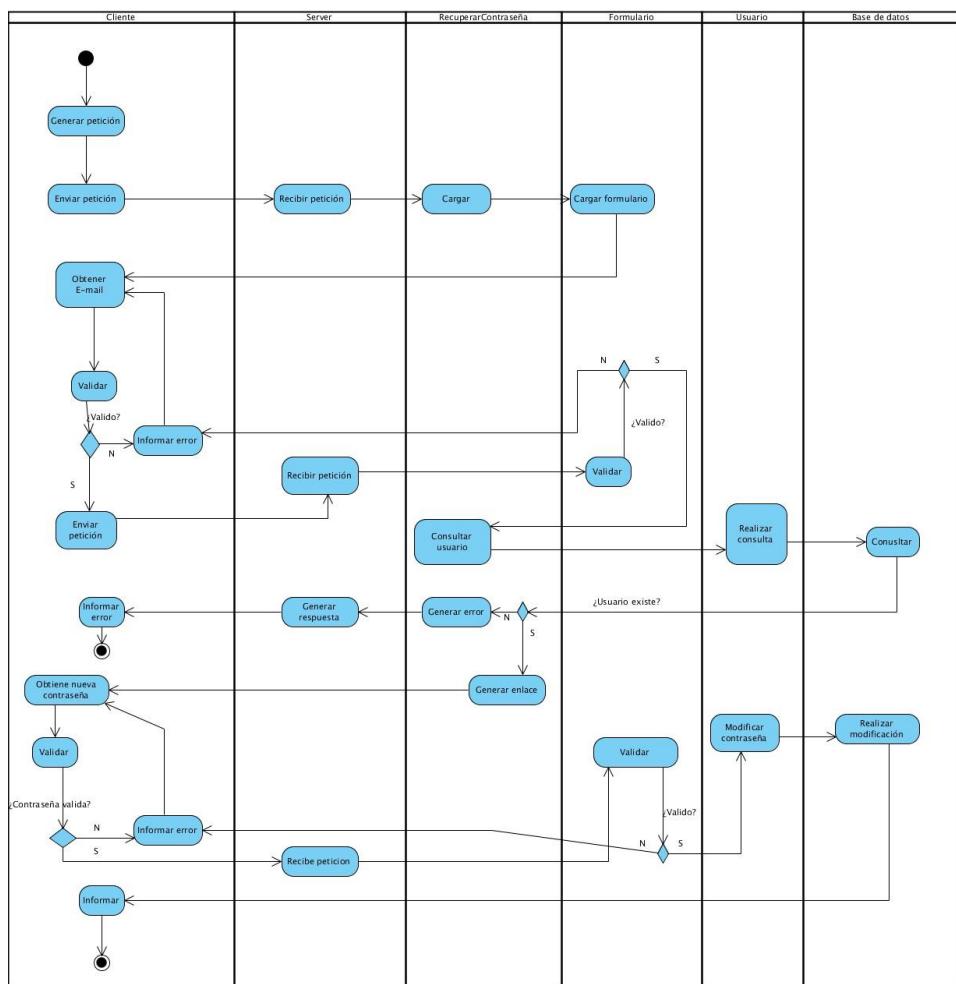


Ilustración 25. Diagrama de actividad Recuperar contraseña

4.2.2 Sistema de establecimientos

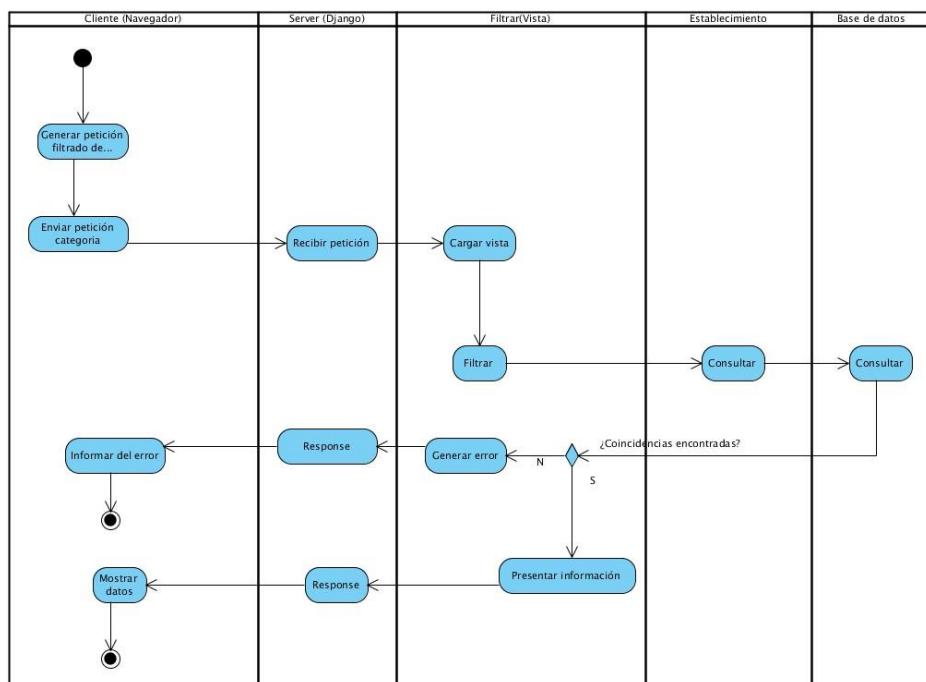


Ilustración 26. Diagrama de actividad Filtrar búsqueda categoría

4.2.3 Sistema recomendador

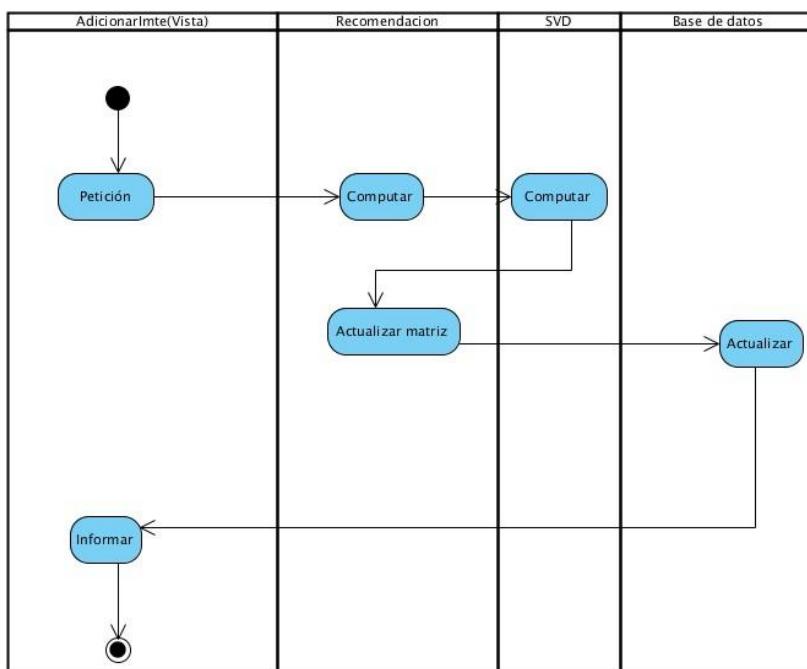


Ilustración 27. Diagrama de actividad Computar

4.3 DIAGRAMAS DE COMUNICACIÓN

A continuación se presenta los diagramas de comunicación para cada sistema, todos los diagramas los encontrará en el anexo virtual Modelamiento del sistema.

4.3.1 Sistema de usuarios

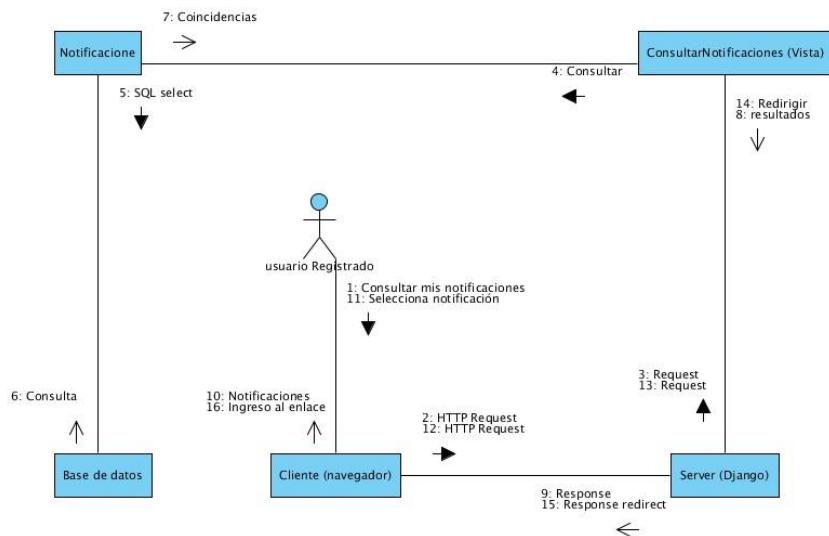


Ilustración 28. Diagrama de comunicación Consultar notificaciones

4.3.2 Sistema de establecimientos

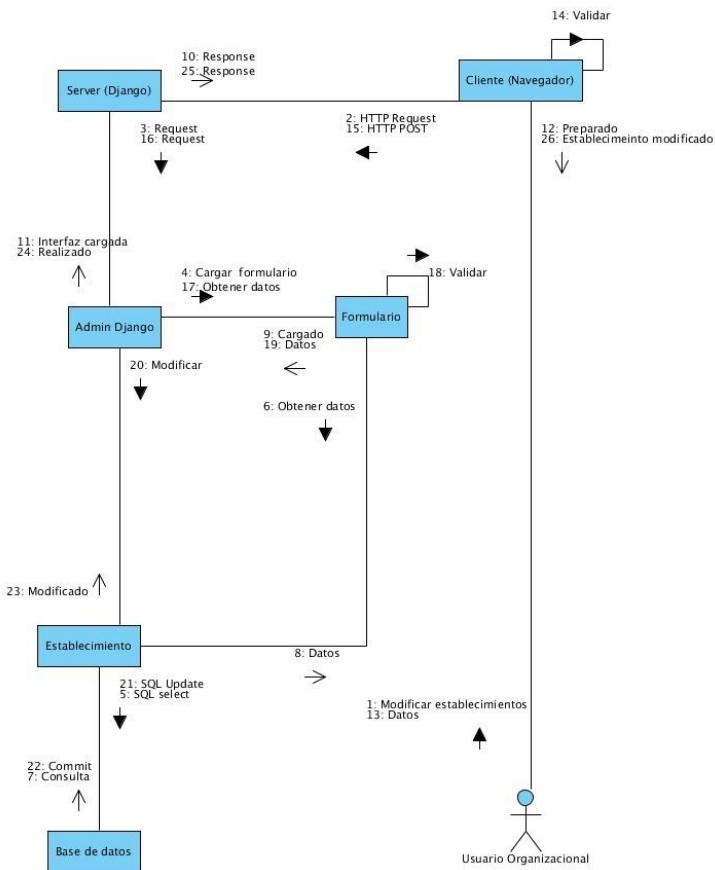


Ilustración 29. Diagrama de comunicación Modificar establecimiento (Administrador)

4.3.3 Sistema recomendador

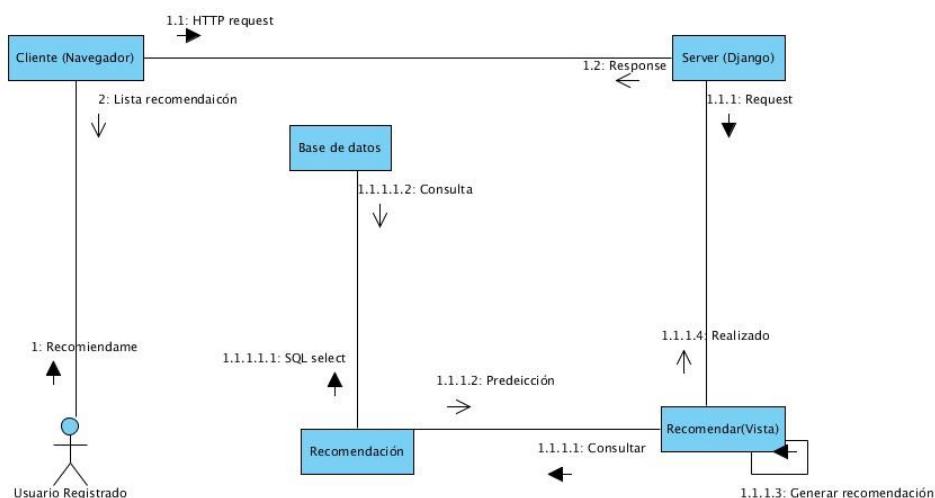


Ilustración 30. Diagrama de comunicación Obtener recomendación

4.4 DIAGRAMA DE ESTADO

A continuación se presentan los diagramas de estados para cada sistema, todos los diagramas los encontrará en el anexo virtual Modelamiento del sistema.

4.4.1 Sistema de usuarios

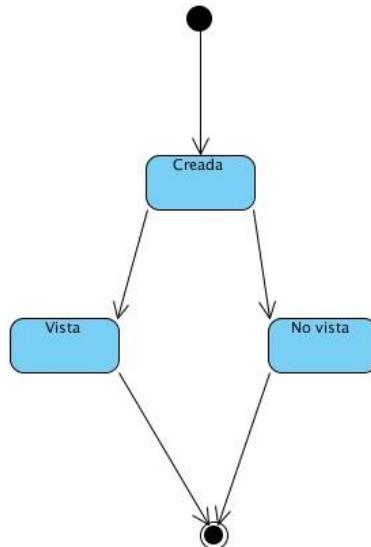


Ilustración 31. Diagrama de estado Notificación

4.4.2 Sistema de establecimientos

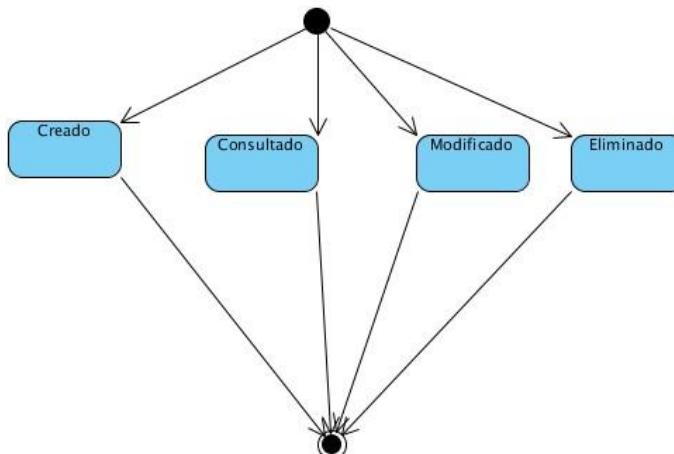


Ilustración 32. Diagrama de estado Establecimiento

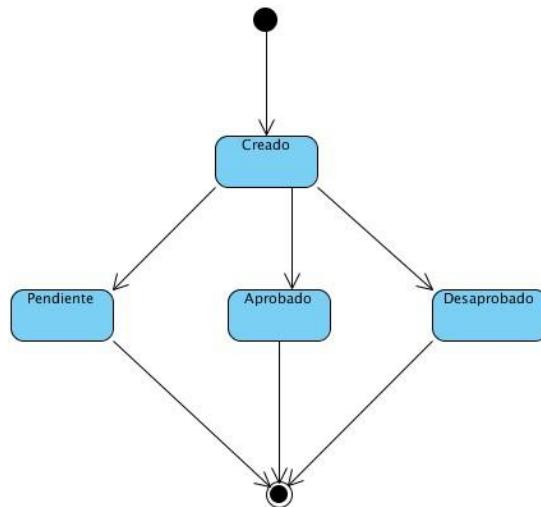


Ilustración 33. Diagrama de estado Solicitud

4.4.3 Sistema recomendador

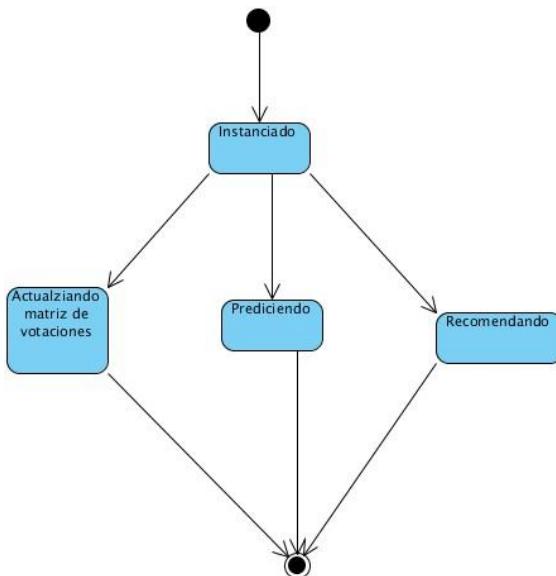


Ilustración 34. Diagrama de estado Recomendación

5 FASE DE DISEÑO

5.1 LISTA PRELIMINAR DE CLASES

A continuación se presenta la lista preliminar de clases para cada sistema.

5.1.1 *Sistema de Usuarios*

Caso de Uso	Clases Vista	Clases Modelo	Clases Formularios	Clases Admin Django
Registrarse	RegistroConvencional	Tipo Usuario(Cuenta)	SignupForm CustomUserCreationForm CustomUserChangeForm	
Registrarse con red Social	RegistroSocial		SingupSocialForm	
Iniciar sesión	LoginViewWithCustomForm		CustomLoginForm	
Cerrar sesión	logout			
Modificar datos de la cuenta	ProfileUpdate		EditAccountForm	
Recuperar contraseña			ChangePasswordForm	
Consultar perfil	PorfilesUsers ProfileUser			
Leer notificaciones	NotificacionesView MarcarTodasNotificacionesLeidas MarcarNotificacionLeida	Notifications	NotificationForm	NotificationsAdmin
Crear usuario				CustomUserAdmin
Modificar usuario				
Eliminar usuario				
Suspender usuario				
Consultar usuarios				

Tabla 17. Lista preliminar de clases del Sistema de usuarios

5.1.2 *Sistema de establecimientos*

Caso de Uso	Clases Vista	Clases Modelo	Clases Formularios	Clases Admin Django
Compartir establecimiento		Establecimiento Categoria SubCategoria	CategoriasFilterForm	

Agregar establecimiento	CrearEstablecimiento		EstablecimientoForm	
Modificar información establecimiento propio	UpdateEstablecimiento			
Adicionar establecimientos Administrador			EstablecimientoAdminForm	EstablecimientoAdmin
Modificar establecimientos Administrador				
Eliminar establecimientos Administrador				
Modificar comentarios		Comentario	ComentarioForm	CommentAdmin
Eliminar comentarios	EliminarComentario			
Agregar imagen a establecimiento	UploadImagenApiView	Imagen	UploadImageForm	ImagenAdmin
Comentar establecimiento	CommentCreateView			
Consultar establecimientos	Establecimientoslist			
Filtrar búsqueda por categoría	EstablecimientosByBoungAutocomplete			
Filtrar búsqueda por cercanía				
Filtrar búsqueda por calificación				
Seleccionar establecimiento	DetalleEstablecimientoView			
Eliminar comentario propio				
Eliminar imagen propia	DeletelImagen			

Eliminar establecimiento propio				
Eliminar imágenes establecimiento propio				
Agregar imagen establecimiento propio				
Consultar establecimientos propios	EstablecimientosPropios			
Modificar comentario propio				
Solicitud eliminación establecimiento no existe	Solicitar	Solicitud TiposSolicitud EstablecimientoTemporal	SolicitudAdminForm SolicitudForm EstablecimientoTemporalForm	Solicitud Admin
Solicitud modificación información				
Solicitud de eliminación de establecimiento por duplicidad				
Solicitud administración establecimiento				

Tabla 18. Lista preliminar de clases del Sistema de establecimientos

5.1.3 Sistema recomendador

Caso de Uso	Clases Vista	Clases Modelo	Clases Formularios	Clases Admin Django
Calificar establecimiento	CalificacionApiView	Voto Score		
Obtener recomendación	RecomendacionView	SVD EstablecimientosRecommender RedisStorage		
Computar				

Tabla 19. Lista preliminar de clases del Sistema recomendador

5.1.4 Lista integrada de clases

Tipo de clase	Clases
Vista	RegistroConvencional RegistroSocial LoginViewWithCustomForm logout ProfileUpdate PorfilesUsers ProfileUser NotificacionesView MarcarTodasNotificacionesLeidas MarcarNotificacionLeida CrearEstablecimiento UpdateEstablecimiento EliminarComentario UploadImagenApiView CommentCreateView Establecimientoslist EstablecimientosByBoung Autocomplete DetalleEstablecimientoView DeleteImagen EstablecimientosPropios Solicitar CalificacionApiView RecomendacionView
Formulario	SignupForm CustomUserCreationForm CustomUserChangeForm SingupSocialForm CustomLoginForm EditAccountForm ChangePasswordForm NotificationForm CategoriasFilterForm EstablecimientoForm EstablecimientoAdminForm ComentarioForm UploadImageForm SolicitudAdminForm SolicitudForm EstablecimientoTemporalForm
Modelo	Tipo Usuario(Cuenta) Notifications Establecimiento Categoria SubCategoria Comentario Imagen Solicitud TiposSolicitud EstablecimientoTemporal

	Voto Score SVD EstablecimientosRecommender RedisStorage
Admin Django	NotificationsAdmin CustomUserAdmin EstablecimientoAdmin CommentAdmin ImagenAdmin SolicitudAdmin

Tabla 20. Lista integrada de clases

5.2 RESPONSABILIDAD DE LAS CLASES

Clase Model	
Operación	Responsable
<code>save()</code>	Tipo
<code>presave</code>	Usuario(Cuenta)
<code>post_save</code>	Notifications
<code>delete()</code>	Establecimiento
<code>__unicode__()</code>	Categoría
<code>__str__()</code>	SubCategoría
<code>Get_absolute_url()</code>	Comentario
<code>__init__()</code>	Imagen
	Solicitud
	TiposSolicitud
	EstablecimientoTemporal
	Voto
	Score

Tabla 21. Responsabilidad de la Clase Model

Clase Forms	
Operación	Responsable
<code>is_valid()</code>	SignupForm
<code>Clean</code>	CustomUserCreationForm
<code>has_changed()</code>	CustomUserChangeForm
<code>Cleaned_data()</code>	SingupSocialForm
<code>as_table()</code>	CustomLoginForm
<code>as_p()</code>	EditAccountForm
<code>as_ul()</code>	ChangePasswordForm
	NotificationForm
	CategoriasFilterForm
	EstablecimientoForm
	EstablecimientoAdminForm
	ComentarioForm

	UploadImageForm SolicitudAdminForm SolicitudForm EstablecimientoTemporalForm
--	---

Tabla 22. Responsabilidad de la Clase Form

Clase View	
Operación	Responsable
Get()	RegistroConvencional
Post()	RegistroSocial
Put()	LoginViewWithCustomForm
Delete()	logout ProfileUpdate PorfilesUsers ProfileUser NotificacionesView MarcarTodasNotificacionesLeidas MarcarNotificacionLeida CrearEstablecimiento UpdateEstablecimiento EliminarComentario UploadImagenApiView CommentCreateView Establecimientoslist EstablecimientosByBoung Autocomplete DetalleEstablecimientoView DeleteImagen EstablecimientosPropios Solicitar CalificacionApiView RecomendacionView

Tabla 23. Responsabilidad de la Clase View

Clase SVD	
Operación	Responsable
Precomputar()	Usuario EstablecimientosRecommender RedisStorage

Tabla 24. Responsabilidad de la Clase SVD

Clase Autocomplete	
Operación	Responsable
Filtrar()	Categoría

	Calificación Establecimiento
--	---------------------------------

Tabla 25. Responsabilidad de la Clase Autocomplete

Clase Solicitud	
Operación	Responsable
Aprobar_eliminacion_duplicidad()	Establecimiento
Aprobar_eliminacion_inexistencia()	Notificación
Aprobar_modificacion()	Usuario
Aprobar_administracion()	

Tabla 26. Responsabilidad de la Clase Solicitud

Clase Notificaciones	
Operación	Responsable
Unread()	Usuario
Read()	Notifications
Mark_as_read()	
Mark_as_unread()	
Mark_all_as_unread()	
Mark_all_as_read()	

Tabla 27. Responsabilidad de la Clase Notificaciones

5.3 MODELO DE ANÁLISIS

Debido a la longitud de los diagramas solo se presentará el diagrama del modelo de Análisis del sistema de recomendador, esto con el fin de facilitar la visibilidad para el lector, el diagrama integrado de modelos análisis lo encontrara en los anexos en formato digital.

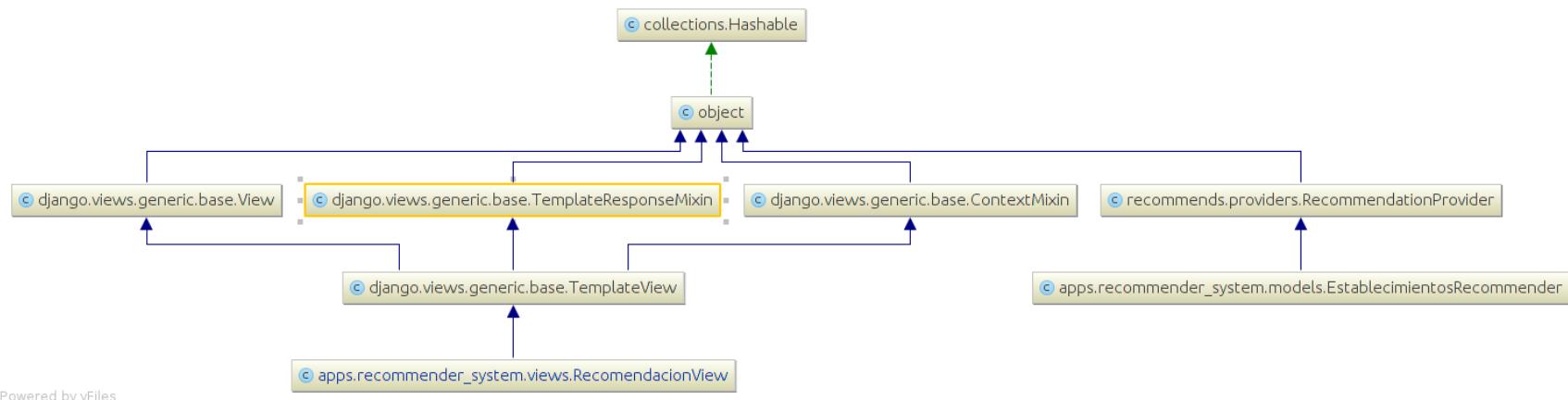


Ilustración 35. Diagrama de clases de Análisis Sistema recomendador

5.4 MODELO LÓGICO

Debido a la longitud de los diagramas solo se presentará el diagrama de modelo Lógico del sistema de recomendador, esto con el fin de facilitar la visibilidad para el lector, el diagrama integrado del modelo Lógico lo encontrara en los anexos en formato digital.

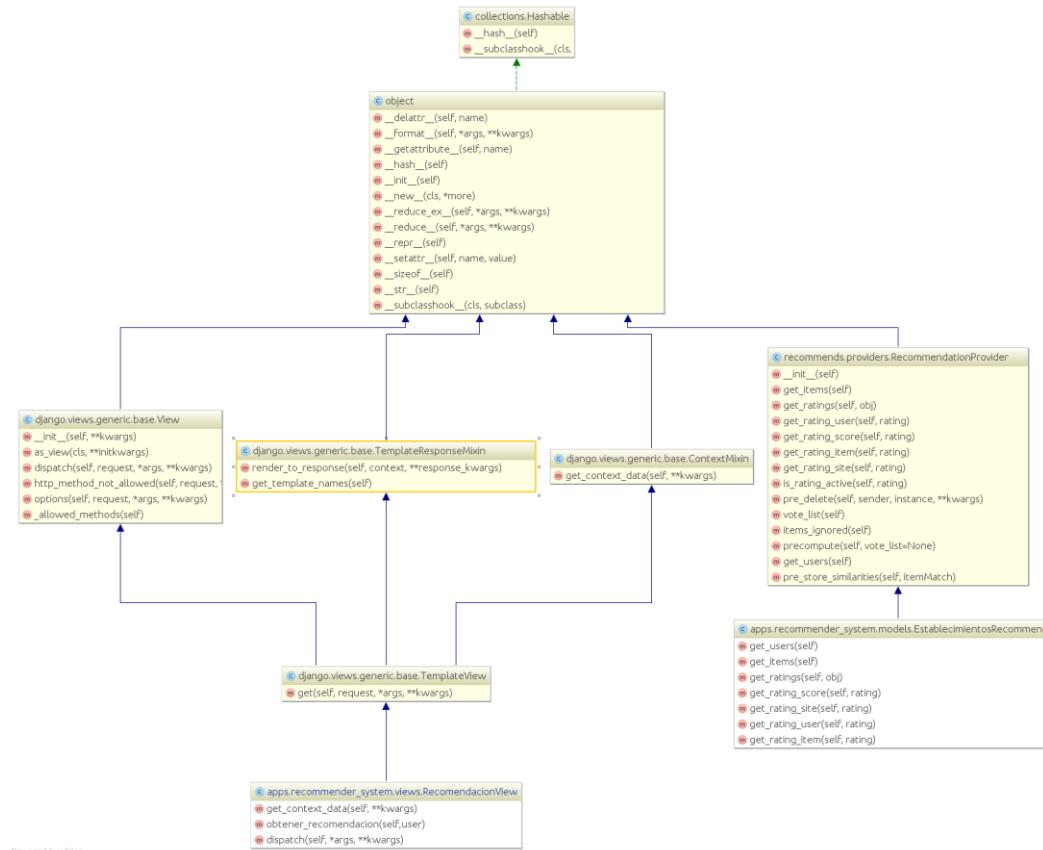


Ilustración 36. Diagrama de clases Lógico Sistema recomendador

5.5 MODELO FÍSICO

Debido a la longitud de los diagramas solo se presentará el diagrama del modelo Físico del sistema de recomendador, esto con el fin de facilitar la visibilidad para el lector, el diagrama integrado del modelo Físico lo encontrara en los anexos en formato digital.

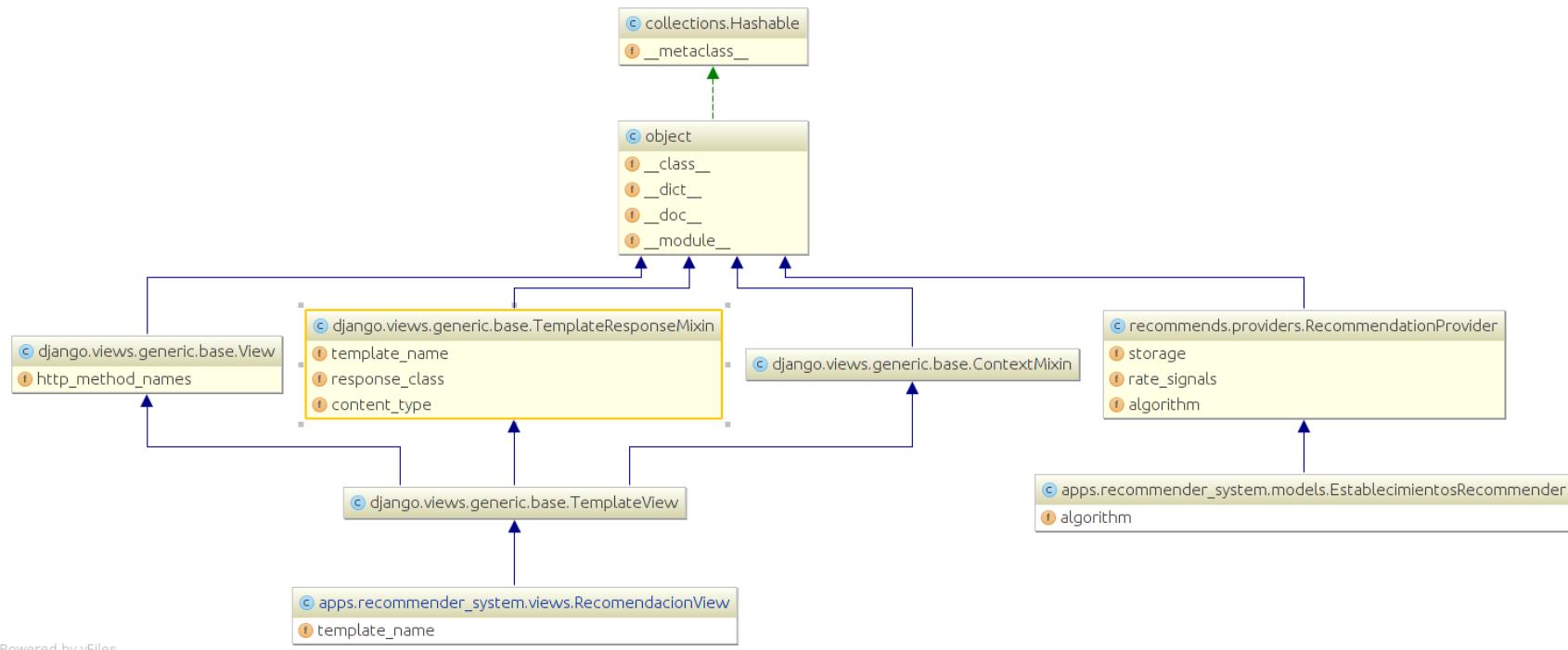


Ilustración 37. Diagrama de clases Físico Sistema recomendador.

5.6 MODELO DE BASE DE DATOS

A continuación se presenta el modelo relacional.

5.6.1 Modelo relacional

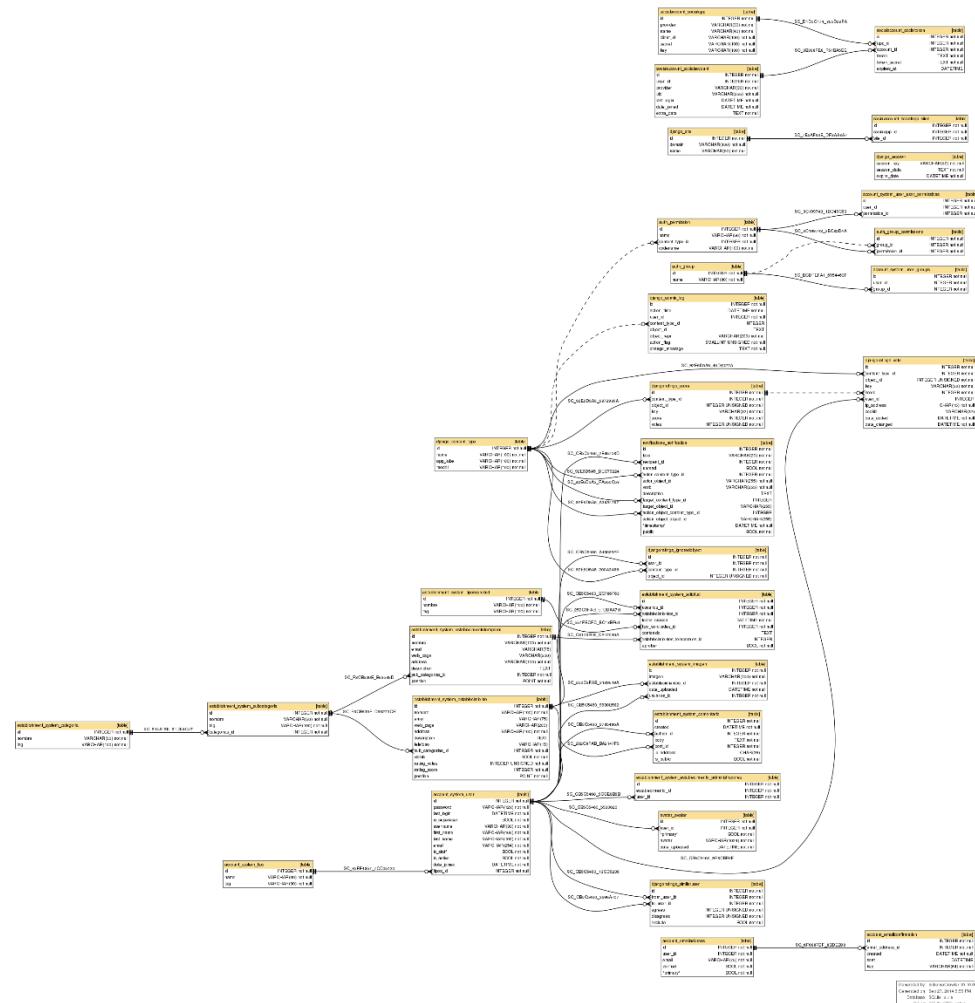


Ilustración 38. Modelo relacional

5.6.2 Modelo de datos de Redis

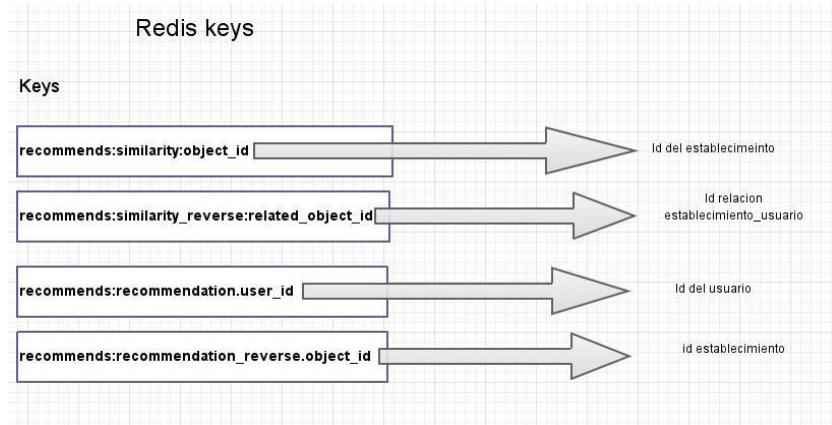


Ilustración 39. Modelo de datos de Redis

5.6.3 Diccionario de datos

Entidad	Descripción	Tipo		Nombre	Propiedad	Llave primaria	Llave foránea	Obligatoria	Varchar	Integer	Position (GEOS)	Datetime	Bpolean
		Fuerte	Débil										
django_admin_log	Tabla que contiene el log del administrador de Django	x		Id	x	x	x	x	x	x		x	
				Action_time			x					x	
				User_id			x		x			x	
				Content_type_id					x				
				Object_id				x	x				
				Object_repr			x	x					
				Action_flag			x		x			x	
				Change_message			x						
auth_permission	Tabla de permisos de usuario	X		Id	x	x	x	x	x	x	x		
				Name			x	x	x	x			
				Content_type_id			x		x	x	x		
				Codename			x	x	x	x			
auth_group_permissions	Relacion permisos grupos	X		Id	x	x	x	x	x	x	x	x	
				Group_id			x		x	x	x	x	
				Permission_id			x	x	x	x	x	x	
auth_group	Grupos de permisos	X		Id	x	x	x	x	x	x	x	x	
				name			x	x	x	x	x	x	
django_content_type		X		Id	x	x	x	x	x	x	x	x	
				Name			x	x	x	x	x	x	

	Type genérico django		App_label		X	X			
			Model		X	X			
django_session	Sesiones de django	X	Sesión_key	x	x	X			
			Sesión_adata		X				
			Expire_date		x				x
django_site	Sitios (Dominios) de django	X	Id	x	X	X			
			Domain		X	X			
			Name		X	x			
account_emailaddresses	Correos cuentas.	X	Id	x	X		X		
			User_id		X		X		
			Email		x	X			x
			Verified						x
			primary						x
account_emailconfirmation	Verificación cuentas	X	Id	X	x	X			
			Email_address_id		X		X		
			Account_emailaddresses		x				
			Created			X			x
			Sent						x
			key		x	x			
socialaccount_socialapp_sites	Relación Sitios sociales con sitios	X	Id	X	X		X		
			Socialapp_id		X		X		
			Site_id		X	x		X	
socialaccount_socialapp	Sitios sociales (Twiter, Facebook, Google+)	X	Id	x	X				
			Provider		X		X		
			Name		X	X			
			Client_id		X	X			
			Secret		X	X			
socialaccount_socialaccount	Datos de usuarios con sitios sociales	X	Key	x	x				
			Id	x	X		X		
			User_id		X		X		
			Provider		X	X			
			Uid		X	X			
			Last_login		X				x
			Date_joined		X				x
socialaccount_socialtoken	Tokens de autenticación de cuentas sociales	X	Extra_data		X				
			Id	X	x		X		
			App_id		X	X	x		
			Account_id		x		X		
			Token			X			
			Token_secret			X			
account_system_tipo	Tipo de usuarios	X	Expires_at						x
			Id	x	X		X		
			Name		X	X			
			tag		x	X			
account_system_user_groups	Relación cuentas con Grupos de usuarios.	X	Id	x	X		X		
			User_id			X		X	
			Group_id		X	x		X	
account_system_user_user_permissions	Relación cuentas con permisos	X	Id	x	x		X		
			User_id			X		X	
			Permission_id		X	x		x	
account_system_user	Usuario	X	Id	x	X		X		
			Password		X	X			
			Last_login		X				x

				Is_superuser		X			x
				Username		X	X		
				First_name		X	X		
				Last_name		X	X		
				Email		X	X		
				Is_staff		X			x
				Is_active		X			x
				Date_joined		X		x	
				Tipos_id		x	X		
djangoratings_vote	Votos	X		Id	X	X	X		
				Content_type_id		X	X	X	
				Object_id		X		X	
				Key		X	X		
				Score		X		X	
				User_id		X		X	
				Ip_address		X	X		
				Cookie			X		
				Data_added		X			x
				Data_changed		X			x
djangoratings_score	Calificaciones	X		Id	X	X	X		
				content_type_id		x	X	X	
				object_id			X		X
				Key		X	X		
				Score		X		X	
				Votes		X		x	
djangoratings_similaruser	Usuarios similares, por votos	X		Id	X	X	X		
				From_user_id		X	X	X	
				To_user_id		x	X		X
				Agress			X		X
				Disagress			X		X
				Exclude		X			x
djangoratings_ignoredobject	Objetos ignorados en la votación	X		Id	X	x	X		
				User_id		X	X	X	
				Content_type_id		X	X		X
				Object_id			X		X
notifications_notification	Notificaciones	X		Id	X	X	X		
				Level			X		
				Recipient_id		X	X		X
				Unread			X		
				Actor_content_type_id		X	X		X
				Actor_object_id			X		
				Verb		X	X		
				Description					
				Target_content_type_id		X			X
				Target_object_id				X	
				Action_object_content_type_id		X			X
				Action_object_object_id				X	
				Timestamp			X		x
				public		X			x
avatar_avatar	Imágenes de perfil	X		Id	X	X	X		
				User_id		X	X		X
				Primary			X		
				Avatar			X	x	

			Date_uploaded		x			X	
establishment_system_categoria	Categorias de establecimientos	X	Id	x	X		X		
			Nombre		X	X			
			Tag		x	x			
establishment_system_subcategoria	Sub_Categorias establecimientos	X	Id	X	X		X		
			Nombre		X	X			
			Tag		X	X			
establishment_system_establecimiento_administradores	Administradores establecimientos	X	Categoria_id	X	x	x			
			Id	X	X		X		
			Establecimiento_id		X		X		
establishment_system_establecimiento	Establecimiento	X	User_id	X	x		X		
			Id	X	X		X		
			Nombre		X	X			
			Email			X			
			Web_page			X			
			Address		X	X			
			Description						
			Teléfono			X			
			Sub_categorias_id	X	X		X		
			Visible		X				x
			Rating_votes		X		X		
			Rating_score		X		X		
establishment_system_imagen	Imágenes	X	position		X			x	
			Id	X	X		X		
			Imagen		X	X			
			Establecimiento_id	X	X		X		
			Date_uploaded		X				X
establishment_system_comentario	Comentarios	X	Usuarios_id	X	X	x			
			Id	X	X		X		
			Created		X				X
			Autor_id	X	X		X		
			Body		X				
			Post_id	X	X		X		
establishment_system_tiposolicitud	Tipo solicitudes		Ip_address			x			
			Is_public		X				X
			Id	X	X		X		
establishment_system_establecimiento_temporal		x	Nombre		X	X			
			Email			X			
			Web_page			X			
			Address		X	X			
			Description						
			Sub_categorias_id	X	X		X		
			position		X			x	
establishment_system_solicitud	Solicitud	x	Id	X	X		X		
			Usuarios_id		X	X		X	
			Establecimiento_id	X	X		X		
			Fecha_creada		X				x
			Tipo_solicitudes_id	X	X		X		
			Contenido						
			establecimientos_temporales_id	x			x		
			aprobar		x				x

Tabla 28. Diccionario de datos

5.7 MODELO DE INTERFAZ

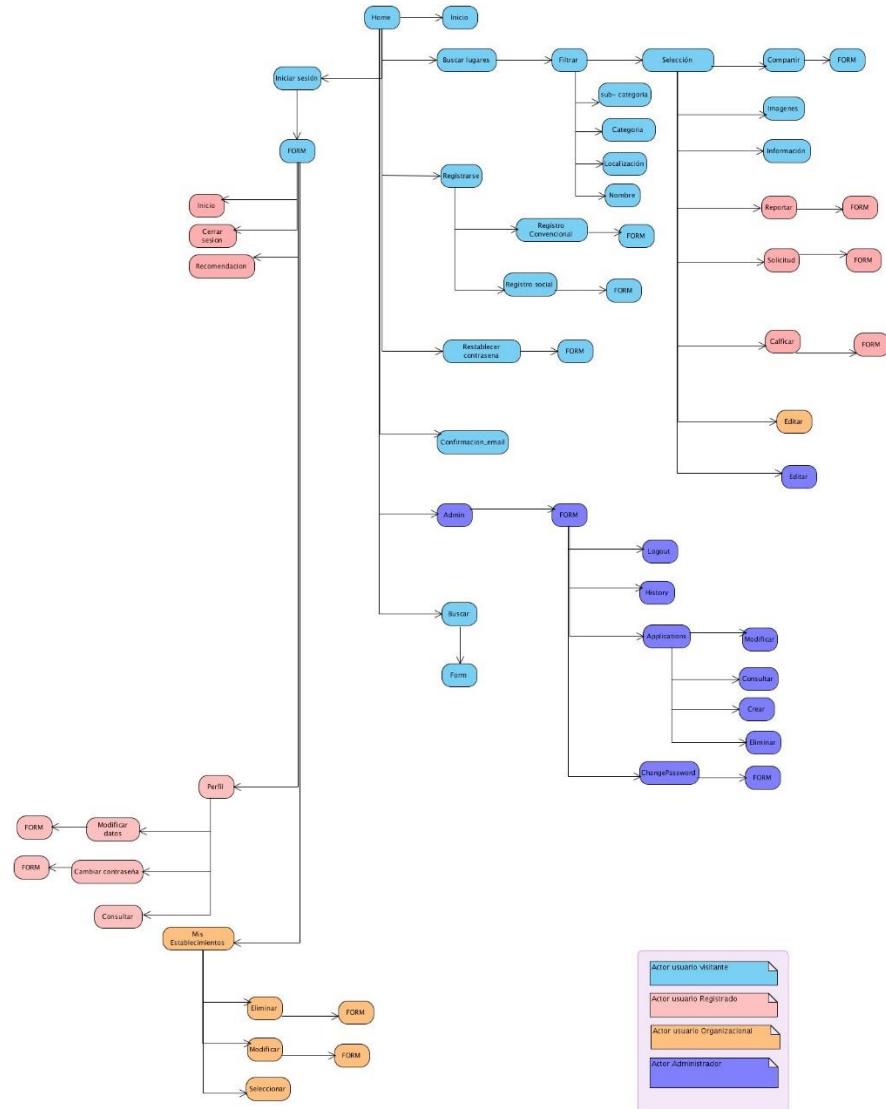


Ilustración 40. Modelo de interfaz

6 FASE DE IMPLEMENTACIÓN

6.1 DIAGRAMA DE COMPONENTES

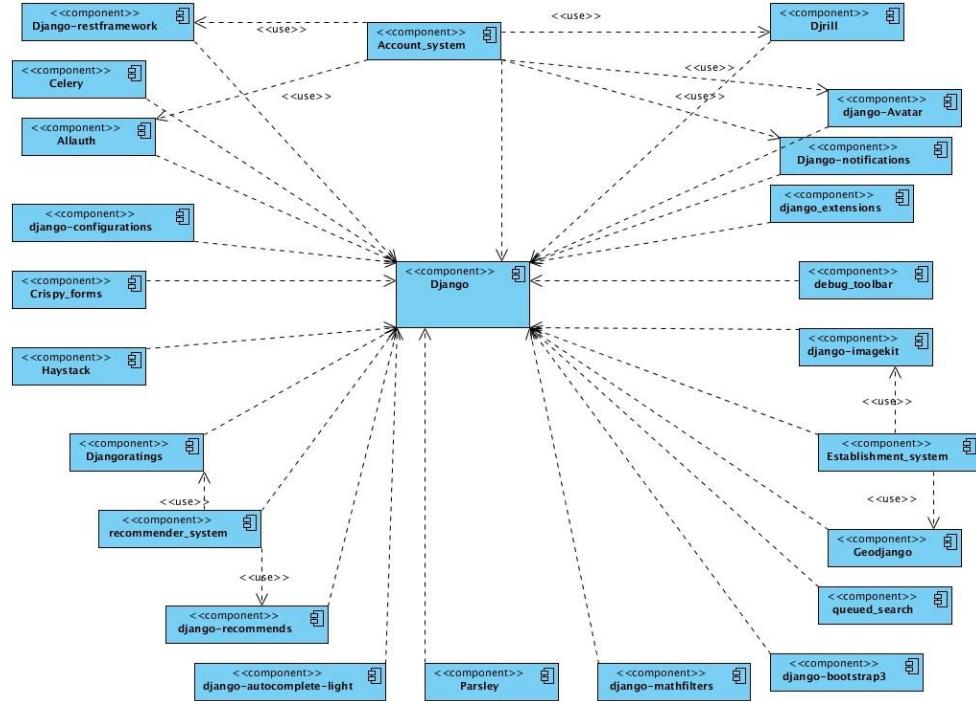


Ilustración 41. Diagrama de componentes

6.2 DIAGRAMA DE PAQUETES

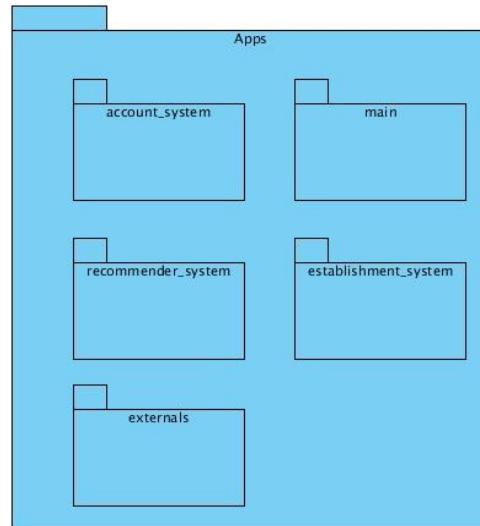


Ilustración 42. Diagrama de paquetes

6.3 DIAGRAMA DE DESPLIEGUE

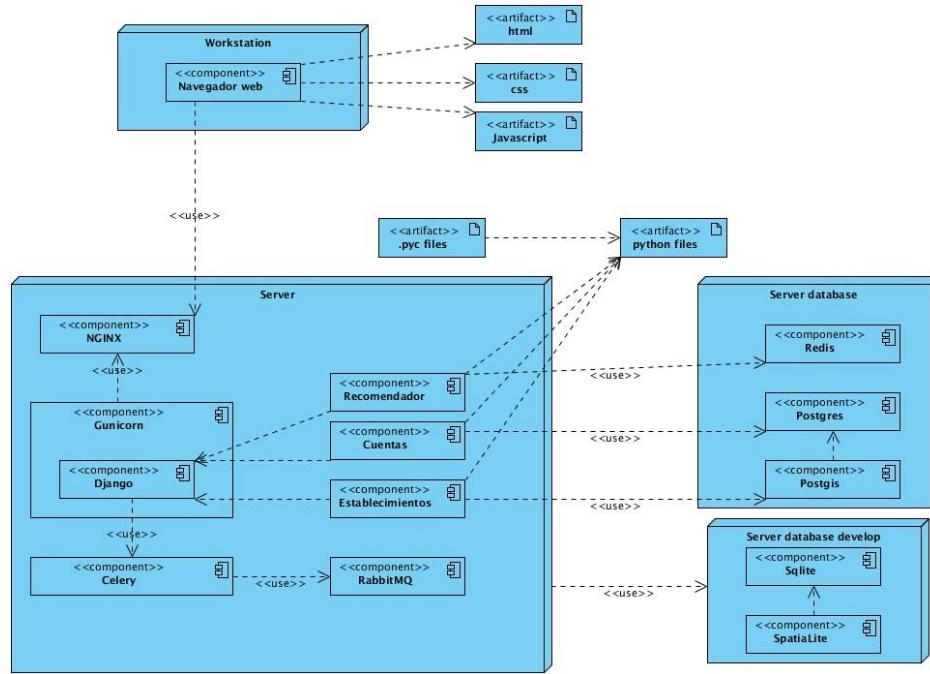


Ilustración 43. Diagrama de despliegue

7 FASE DE PRUEBAS

7.1 Pruebas de integración

7.1.1 Sistema de usuarios

Sistema de usuarios			
Autor	Camilo Ramírez		
Inicio	10:00 am		
Fin	12:15 m		
Concepto	Revisar el correcto funcionamiento del sistema de usuarios.		
Acción	Usuario	Resultado esperado	Estado
Registro convencional	No registrado	Crear un nuevo usuario mediante el registro convencional.	Finalizado
Registro social	No registrado	Crear un nuevo usuario mediante el registro con redes sociales.	Fallo 1
Iniciar sesión	No registrado	Iniciar sesión con un usuario previamente registrado y verificado.	Finalizado
Restablecer contraseña	No registrado	Restablecer contraseña de un usuario previamente registrado.	Finalizado
Cerrar sesión	registrado	Finalizar sesión, no permitir ingresar a enlaces que contenga información de una sesión finalizada.	Fallo 3
Modificar cuenta	registrado	Modificar los datos del usuario.	Fallo 4
Consultar notificaciones	registrado	Consultar sus notificaciones activas.	Finalizado
Gestionar comentarios.	registrado	Crear y eliminar un comentario realizado por el mismo usuario.	Fallo 5
Consultar perfil	registrado	Consultar perfil del propio usuario.	Fallo 6
Gestión notificaciones	Administrador	Crear, consultar, editar y eliminar notificaciones de cualquier usuario.	Finalizado

Gestión de usuarios	Administrador	Crear, consultar, editar, eliminar y suspender cualquier tipo de usuario.	Finalizado
Gestión de comentarios	administrador	Crear, consultar, editar y eliminar comentarios creado por cualquier usuario.	Finalizado

Tabla 29 . Pruebas sistema de usuarios

Fallos			
Seguimiento	Prioridad	Descripción del fallo	Corregido
Fallo 1	Media	EL formulario de nombres y apellidos en el registro social permite ingresar números. Este fallo también se genera en el formulario de actualización de perfil.	SI
Fallo 3	Baja	Si se dirige a un enlace protegido con contraseña se abrirá la interfaz de iniciar sesión, al iniciar sesión satisfactoriamente se direcciona a un enlace roto y lanzando un error 404.	SI
Fallo 3	Alta	Se genera un error de configuración al momento de registrarse mediante Facebook genera.	SI
Fallo 4	Baja	Cuando se modifican los datos del perfil o la contraseña de usuario no se informa si se efectuaron los cambios.	SI
Fallo 5	Media	Permite crear comentarios con varios espacios en blanco de manera seguida.	SI
Fallo 6	Alta	Para el usuario Administrador se presenta un error al eliminar un comentario creado por el mismo.	SI

Tabla 30. Fallos Sistema de usuarios

7.1.2 Sistema de establecimientos

Sistema de establecimientos	
Autor	Camilo Ramírez
Inicio	12:15 m
Fin	2:30 pm

Concepto	Revisar el correcto funcionamiento del sistema de establecimientos.		
Acción	Usuario	Resultado esperado	Estado
Consulta de establecimientos	Usuario visitante	Filtrar los establecimientos por categoría, sub categoría o localización.	Finalizado
Compartir establecimiento	Usuario visitante	Compartir el enlace del establecimiento por redes sociales, Facebook, Twitter o Google+.	Finalizado
Solicitud de administración	Usuario registrado	Crear una nueva solicitud de administración por el usuario creado.	Finalizado
Solicitud de modificación.	Usuario registrado	Crear una nueva solicitud de modificación.	Fallo 1
Solicitud eliminación duplicidad	Usuario registrado	Crear una nueva solicitud de eliminación por duplicidad.	Finalizado
Solicitud eliminación inexistencia.	Usuario registrado	Crear una nueva solicitud de eliminación por inexistencia.	Finalizado
Gestión de imágenes	Usuario registrado	Crear y eliminar imágenes subidas por el propio usuario, solo debe permitir subir un máximo de 3 por usuario en establecimiento y no más de 8 por establecimiento.	Fallo 2
Crear establecimiento	Usuario registrado	Crear un establecimiento.	Finalizado
Gestión de establecimientos	Usuario organizacional	Consultar, modificar y eliminar establecimientos pertenecientes al usuario.	Fallo 3
Gestión de imágenes	Usuario organizacional	Crear, consultar y eliminar imágenes pertenecientes a sus establecimientos.	Finalizado
Gestión de establecimientos	Usuario administrador	Crear, consultar, modificar y eliminar establecimientos.	Finalizado
Gestión de imágenes.	Usuario administrador	Crear, consultar, modificar y eliminar.	Finalizado
Aprobar solicitud administración	Usuario administrador	Aprobar solicitud de administración.	Finalizado

Aprobar solicitud modificación	Usuario administrador	Aprobar solicitud de modificación.	Finalizado
Aprobar solicitud de eliminación	Usuario administrador	Aprobar solicitud de eliminación..	Finalizado

Tabla 31. Pruebas sistema de establecimientos

Fallos			
Seguimiento	Prioridad	Descripción del fallo	Corregido
Fallo 1	Baja	Cundo se genera una solicitud de modificación si esta contiene un nombre de algún establecimiento ya existe y posteriormente es aprobada se emite un error de inserción.	SI
Fallo 2	Baja	Si un usuario sube tres imágenes en distintos lados el sistema no permite que suba más imágenes en otros establecimientos	SI
Falle 3	Alta	Al crear un establecimiento y encontrar y enviar errores nos e guarda la posición del establecimiento.	SI

Tabla 32. Fallos sistema de establecimientos

7.1.3 Sistema recomendador

Sistema recomendador			
Autor	Camilo Ramírez		
Inicio	2:30 pm		
Fin	3:30 pm		
Concepto	Revisar el correcto funcionamiento del sistema recomendador.		
Acción	Usuario	Resultado esperado	Estado

Calificar	Registrado	Mensaje de que la calificación se realizó.	Finalizado
Cambiar calificación	Registrado	Mensaje de que la calificación se modificó.	Finalizado
Obtener recomendación	Registrado	Obtener los 10 mejores establecimientos según los criterios de búsqueda del usuario, si el número de establecimientos recomendados es menor a 10 añadir los mejores establecimientos hasta completar el total de 10.	Fallo 1

Tabla 33. Pruebas sistema recomendador

Fallos			
Seguimiento	Prioridad	Descripción del fallo	Corregido
Fallo 1	Alta	Cuando no se obtiene recomendaciones el sistema recomienda los 10 mejores establecimientos, el problema es que si alguno de ellos ya fue puntuado por el usuario aun así lo recomienda.	SI

Tabla 34. Fallos sistema recomendador

7.2 Pruebas de sistema

Pruebas de sistema			
Autor	Camilo Ramírez		
Inicio	5:30 pm		
Fin	9:30 pm		
Concepto	Realizar las pruebas del sistema.		
Acción	Usuario	Resultado esperado	Estado
Eliminar usuarios que sean propietarios de un establecimiento	Administrador	Eliminar los usuarios del campo de administrador de un establecimiento.	Finalizado

Eliminar usuario que calificó un establecimiento.	Administrador	Eliminar el usuario y la calificación dada al establecimiento, así como sus recomendaciones.	Finalizado
Eliminar usuario que creó un comentario.	Administrador	Eliminar el comentario y el usuario.	Fallo 1
Eliminar un usuario administrador.	Administrador	Eliminar el usuario.	Finalizado
Eliminar comentario realizado por otro administrador	Administrador	Eliminar un comentario realizado por otro administrador.	Finalizado
Eliminar un establecimiento que previamente haya sido calificado.	Administrador	Eliminar el establecimiento y de la misma manera todos las calificaciones.	Fallo 2
Eliminar establecimiento que tenga imágenes	Administrador	Eliminar el establecimiento junto con las imágenes.	Finalizado
Eliminar establecimiento que pertenezca a un usuario.	Administrador	Eliminar establecimiento que pertenezca a un usuario propietario.	Finalizado
Eliminar una solicitud activa.	Administrador	Eliminar la solicitud y también los establecimientos temporales (Si tiene)	Finalizado
Eliminar imágenes pro parte de un propietario.	Organizacional	Eliminar las imágenes pertenecientes a un usuario propietario.	Finalizado
Editar establecimiento de un propietario.	Administrador	Editar la información de un establecimiento perteneciente a un usuario.	Finalizado

Tabla 35. Pruebas de sistema

Fallos			
Seguimiento	Prioridad	Descripción del fallo	Corregido

Fallo 1	Alta	Al eliminar el usuario el comentario no se elimina creando problemas de integridad de la bases de datos.	SI
Fallo 2	Alta	Al eliminar el establecimiento no se eliminan las calificaciones generando problemas al momento de recomendar estos establecimientos a los usuarios.	Si

Tabla 36. Fallos de sistema

7.3 Pruebas del algoritmo de filtrado colaborativo SVD

A continuación presentaremos las pruebas realizadas al algoritmo SVD en su implementación de Pyrecsys.

Para la realización de la prueba es necesario contar con una cantidad considerable de datos confiables, para recolectar esta información se necesita conocer las preferencias de gran cantidad de usuarios respecto a distintos ítems, tal información solo podrá ser proporcionada por los mismos usuarios. Esto supone un problema ya que es muy difícil obtener esta información ya que requiere de tiempo, usuarios y esfuerzo por los anteriores para proporcionar esa información en los momentos iniciales del sistema donde no proporcionará buenas recomendaciones.

Debido a la dificultad anterior no contaremos con datos proporcionados por el portal web de recomendación realizado en este proyecto. Para superar el inconveniente usaremos una de las bases de datos proporcionada por el grupo de investigación GropuLens¹⁰ de la universidad de Minnesota, tal base de datos *Movielens1M* consiste de aproximadamente 1 millón de votos proporcionados por 6040 usuarios para 3900 películas, con puntajes entre 1 y 5, siendo una base de datos que se acopla muy bien para el uso con nuestro algoritmo SVD.

7.3.1 ERROR MEDIO ABSOLUTO (MAE)

Esta prueba consistirá en definir la fiabilidad de las recomendaciones, es decir conocer que tan exacta serán sus predicciones, para ello usaremos el *dateset* de *Movielens* donde el 80% será usado como datos de entrenamiento (*Trainingtest*) y el resto como datos de prueba (*Datetest*), la idea es predecir los datos del conjunto de pruebas considerando únicamente los conjuntos de entrenamiento como datos

¹⁰ <http://grouplens.org/datasets/movielens/>

de entrada. Por último se comparan los datos obtenidos con los reales, para esto usaremos una métrica muy conocida en los sistemas de recomendación llamada *MAE* (*Mean Absolute error*) que mide la desviación media entre el valor de predicción real que el usuario asignó al ítem.

La fórmula de cálculo para el MAE es la siguiente:

$$\overline{|E|} = \frac{\sum_{i=1}^N |p_i - r_i|}{N}$$

Donde:

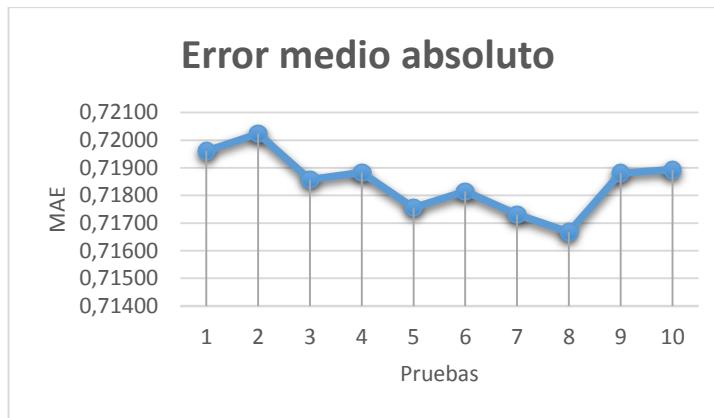
- $|E|$ = Valor absoluto promedio del error
- p_i = Puntaje de predicción del algoritmo sobre el ítem i
- r_i = Puntaje real del usuario sobre el ítem i
- N = Total de elementos considerados para el cálculo de MAE

A partir de esta métrica cada elemento del conjunto de pruebas es tratado de igual manera, es decir que se le da el mismo peso al error cometido en calcular la predicción de un elemento donde su valor real está por encima o por debajo del promedio de la escala de valores.

La prueba se realizó 10 veces con diferentes datos de entrenamiento y de pruebas, con un valor de $k=100$, cada conjunto de entrenamiento tenía 800167 tuplas y de pruebas tenía 200042 tuplas. Los resultados obtenidos fueron los siguientes:

Iteración	Densidad de la matriz	MAE
1	3,5960	0,71961
2	3,6019	0,72023
3	3,6029	0,71857
4	3,6039	0,71883
5	3,6078	0,71756
6	3,5980	0,71815
7	3,6058	0,71731

8	3,6048	0,71667
9	3,5951	0,71881
10	3,6058	0,71893



Por lo tanto el MAE global es de **0,71847**.

7.3.2 *Capacidad de recomendación Coverage [50]*

El *coverage* (Cobertura) es simplemente el porcentaje de elementos para los cuales el sistema podrá generar una predicción. Para nuestro caso se puede decir que SVD tiene una capacidad de recomendación del 100% siempre y cuando obtengamos las matrices de factores de usuarios e ítems, siendo una de las más grandes ventajas de este tipo de algoritmos.

7.3.3 *Tiempos de ejecución*

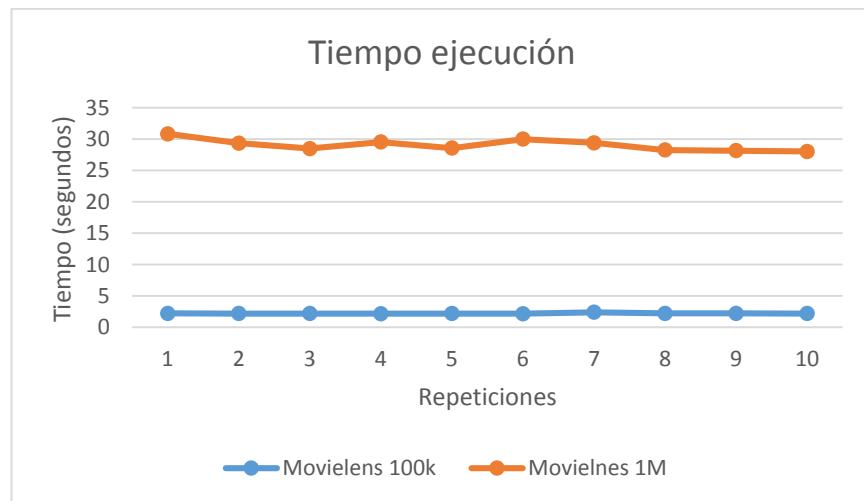
Para esta prueba se usó un computador portátil con Ubuntu 14.04 y con el siguiente hardware:

- Procesador Intel Core i7 36-100M de 2.3GH
- Memoria RAM de 8 GB

En este caso se usaron varias bases de datos de *Movielens*, la de 100k y la de 1M, los resultados obtenidos para la función de entrenamiento fueron los siguientes:

#	Movielens 100k (s)	Movielnes 1M (s)
1	2,216796875	28,60504699
2	2,195018053	27,1582489

3	2,198416948	26,28635097
4	2,183682919	27,34268999
5	2,200455189	26,35942101
6	2,174726009	27,82125902
7	2,405908108	26,98850918
8	2,21942091	26,03176188
9	2,227161884	25,92684603
10	2,199163914	25,81643796



El promedio de tiempo de ejecución de entrenamiento para la base de datos de *Movielens100k* fue de aproximadamente **2,2 segundos** Mientras que con la base de datos de *Movielens1M* fue de **26,8 segundos** aproximadamente.

CONCLUSIONES

- El uso de cuestionarios para conocer los requerimientos y necesidades de propietarios de microempresas, facilitó el planteamiento del problema y la justificación.
- Se analizaron los algoritmos, métodos y técnicas más usados en los sistemas de recomendación, especialmente en el filtrado colaborativo, de igual forma las métricas más utilizadas para medir su efectividad y rendimiento.
- Se analizaron proyectos y soluciones de referencia en el área de los sistemas de recomendación con el fin de conocer los diferentes enfoques y aplicaciones de los sistemas de recomendación.
- Se aplicó el algoritmo SVD (Singular Value Descomposition) de filtrado colaborativo con las mejoras planteadas por Simon Funk, en la recomendación de establecimientos comerciales lo cual permitió la generar recomendaciones de establecimiento comerciales con una buena calidad a los usuarios del sistema
- Se logró obtener las preferencias de los usuarios mediante captura explicitas en rango de puntuaciones, con el fin de representar el grado de satisfacción de cada usuario.
- Se realizaron pruebas de integración y de sistema para el portal web, de igual manera se realizaron pruebas de efectividad (MAE), tiempo de ejecución, y cobertura (Coverage) en el algoritmo de filtrado colaborativo SVD, lo que evidenció el poder de cómputo y la complejidad del algoritmo SVD aplicado al filtrado colaborativo.
- La separación por módulos del sistema organiza y facilita las diferentes fases de desarrollo, logrando así una independencia entre implementaciones en el portal Web.

Bibliografía

- [1] C. Eridon, «The History of Marketing: An Exhaustive Timeline [Infografía],» [En línea]. Disponible: <http://blog.hubspot.com/blog/tabid/6307/bid/31278/The-History-of-Marketing-An-Exhaustive-Timeline-INFOGRAPHIC.aspx>.
- [2] «Internet, plataforma de crecimiento de las empresas»,» El Espectador, [En línea]. Disponible: <http://elespectador.com/tecnologia/empresas/articulo-409568-internet-plataforma-de-crecimiento-de-empresas>. [Último acceso: 2013 mar 27].
- [3] P. Marketing, «La publicidad móvil comienza a ganar peso en los planes de marketing de las empresas,» [En línea]. Disponible: <http://www.puromarketing.com/21/10517/publicidad-movil-comienza-ganar-peso-planos-marketing-empresas.html..> [Último acceso: 2013 mar 27].
- [4] ITespresso.es, «¿Demasiada información o mucha información desestructurada?,» [En línea]. Disponible: <http://www.itespresso.es/%c2%bfdemasiada-informacion-o-mucha-informacion-desestructurada-37313.html..> [Último acceso: 2013 mar 27].
- [5] «El arte de presentar » La mucha información confunde más que orienta»,» [En línea]. Disponible: [http://www.elartedepresentar.com/2008/11/la-mucha-informacion-confunde-mas-que-orienta/.](http://www.elartedepresentar.com/2008/11/la-mucha-informacion-confunde-mas-que-orienta/>.) [Último acceso: 27 mar 2013].
- [6] V. F. López, «Técnicas eficientes para la recomendación de productos basadas en filtrado colaborativo,» Universidad de la Coruña, 2013.
- [7] R. M. García, «SVD APLICADO A SISTEMAS DE RECOMENDACIÓN BASADOS EN FILTRADO COLABORATIVO,» Universidad Politécnica de Madrid, Madrid, 2012-2013.
- [8] R. B. y. V. Carus, *Internet y comercio electrónico*, 2000.
- [9] A. d. C. B. y. C. d. C. d. Bogotá, *Censo empresarial de Ciudad Bolívar*, Bogotá , Colombia: Alcaldía de Ciudad Bolívar.
- [10] C. d. C. d. Bogotá, « Estadísticas empresas registradas en Bogotá y 59 municipios de Cundinamarca,» [En línea]. Disponible: <http://wwwccb.org.co/contenido/contenido.aspx?catID=94&conID=10510>. [Último acceso: 22 Abril 2013].

- [11] «Estadísticas del número de empresas en Bogotá y 59 municipios de Cundinamarca,» [En línea]. Disponible: http://www.empresario.com.co/recursos/page_flip/CCB/2012/estadisticas_empresas_bogota_59municipios/#/1/. [Último acceso: 8 Mayo 2013].
- [12] S. d. i. S. S. d. H. y. H. v. H. I. n. A. mayor de Bogotá, *LECTURA DE REALIDADES Y NECESIDADES TERRITORIO PAS Perdomo, Arborizadora, San Francisco LOCALIDAD 19 CIUDAD BOLÍVAR GESTIÓN SOCIAL INTEGRAL*, Bogotá, Colombia: Alcaldía Mayor de Bogotá.
- [13] Canonical, «Ubuntu,» [En línea]. Disponible: <http://www.ubuntu.com/>. [Último acceso: 2014 11 1].
- [14] Python, «Python,» [En línea]. Disponible: <https://www.python.org/about/>. [Último acceso: 2014 08 1].
- [15] L. Torvalds, «Git,» [En línea]. Disponible: <http://git-scm.com/>. [Último acceso: 2014 10 9].
- [16] D. S. Foundation, «Django Project,» [En línea]. Disponible: <https://www.djangoproject.com/>. [Último acceso: 2014 2 23].
- [17] Twitter, «Twitter Bootstrap,» [En línea]. Disponible: <http://getbootstrap.com/>. [Último acceso: 2014 05 15].
- [18] T. j. Foundation, «jQuery write less, do more,» [En línea]. Disponible: <http://jquery.com/>. [Último acceso: 2014 4 18].
- [19] A. Solem, «Celery project,» [En línea]. Disponible: <http://www.celeryproject.org/>. [Último acceso: 2014 10 10].
- [20] R. T. Ltd, «Rabbitmq,» [En línea]. Disponible: <http://www.rabbitmq.com/>. [Último acceso: 2014 10 14].
- [21] Nginx, «NGINX,» [En línea]. Disponible: <http://wiki.nginx.org/Main>. [Último acceso: 2014 2 8].
- [22] Citrusbyte, «Redis,» [En línea]. Disponible: <http://redis.io/>. [Último acceso: 2014 24 7].
- [23] T. P. G. D. Group, «Postgresql,» [En línea]. Disponible: <http://www.postgresql.org/>. [Último acceso: 2014 13 11].

- [24] PostGis, «Postgis,» [En línea]. Disponible: <http://postgis.net/support>. [Último acceso: 2014 10 14].
- [25] D. R. Hipp, «Sqlite,» [En línea]. Disponible: <http://www.sqlite.org/about.html>. [Último acceso: 2014 9 14].
- [26] A. Furieri, «SpatiaLite,» [En línea]. Disponible: <https://www.gaia-gis.it/fossil/libspatialite/index>. [Último acceso: 2014 10 14].
- [27] Google, «Google maps,» [En línea]. Disponible: <https://www.google.com/maps/about/>.
- [28] A. P. A. Rodríguez, *Sistema de recomendación colaborativo basado en algoritmos de filtrado mejorados*, España: Escuela Politécnica Superior de Jaén.
- [29] C. R. M. S. V. S. y. C. d. C. Enrique García Salcines, *Sistema recomendador colaborativo usando minería de datos distribuida para la mejora continua de cursos e-learning*, IEEE-RITA, 2008.
- [30] «Sistemas de recomendaciones: herramientas para el filtrado de información en Internet - Hipertext,» [En línea]. Disponible: <http://www.upf.edu/hipertextnet/numero-2/recomendacion.html>. [Último acceso: 10 abril 2013].
- [31] D. L. M. Lope, *Un nuevo modelo ponderado para Sistemas de Recomendación Basados en Contenido con medidas de contingencia y entro*, España: Universidad de Jaén, 2012.
- [32] J. M. F.-L. J. F. H. y. M. A. R.-M. L. M. de Campos, *USO DE CONOCIMIENTO ESTRUCTURADO EN UN SISTEMA DE RECOMENDACIÓN BASADO EN CONTENIDO*.
- [33] G. Adomavicius y A. Tuzhilin, *Toward the Next Generation of Recommender Systems: A Survey of the State-of-the-Art and Possible Extension*, .IEEE Transactions on Knowledge and Data Engineering .
- [34] «¿Que son los Portales en Internet?,» [En línea]. Disponible: <http://www.informaticamilenium.com.mx/paginas/mn/articulo25.html>. [Último acceso: 11 Mayo 2013].
- [35] J. C. G. Góme, *Portales de internet: concepto, tipología básica y desarrollo*, Universidad de Murcia.
- [36] A. R. K. P. M. y. L. E. C. P. C. Abad, *Sistema de Recomendación de Películas*, Guayaquil-Ecuador: Escuela Superior Politécnica del Litoral, 2010.

- [37] A. A. Ibáñez, «Sistema de recomendación de restaurantes georreferenciados,» Universidad de Jaén, Escuela Politécnica Superior de Jaén, Jaén - España, 2008.
- [38] V. C. Busquet, Sistema de recomendación personalizada de contenido, 2008.
- [39] D. J. F. B. Rivera, *Prototipo Web de un sistema de Recomendación de recursos Docentes*, Jaén , España: Universidad de Jaén, 2011.
- [40] Qype, «Sobre Qype,» [En línea]. Disponible: <http://www.qype.es/impressum..> [Último acceso: 12 abril 2013].
- [41] I. e. I. Wikipedia, «Qype,» [En línea]. [Último acceso: 06 abril 2013].
- [42] Yelp, «Preguntas frecuentes | Yelp»,» [En línea]. Disponible: http://www.yelp.es/faq#what_is_yelp. [Último acceso: 12 Abril 2013].
- [43] Yelp, «Sobre Yelp,» [En línea]. Disponible: <http://www.yelp.es/about>. [Último acceso: 12 Abril 2013].
- [44] Itespresso, «Operación review: Yelp se hace con Qype por 50 millones de dólares,» [En línea]. Disponible: <http://www.itespresso.es/yelp-compra-qype-review-69896.html..> [Último acceso: 12 Abril 2013].
- [45] Guru, «Guru – Comunidad de recomendaciones de Colombia,» [En línea]. Disponible: <http://www.ciudadguru.com.co/guru..> [Último acceso: 29 Abril 2013].
- [46] S. M. G. Nieto, «Filtrado Colaborativo y Sistemas de recomendación,» Universidad Carlos III de Madrid,Getafe, Leganés y Colmenarejo España, [En línea]. Disponible: <http://www.it.uc3m.es/jvillena/irc/practicas/06-07/31.pdf..> [Último acceso: 13 Abril 2013].
- [47] C. d. C. d. Bogotá, «Instructivos, Paso a paso. Establecimiento Comercial.,» 1 10 2014. [En línea]. Disponible: http://recursos.ccb.org.co/ccb/instructivos/paso_paso_empresa_ccb/Establecimiento_Comercio.html.
- [48] J. F. Vega, «Qué significa backend y frontend en el diseño web,» [En línea]. Disponible: <http://www.cristalab.com/blog/que-significa-backend-y-frontend-en-el-diseno-web-c106224/>. [Último acceso: 11 Mayo 2013].

- [49] Alegsa, «Alegsa.com.ar,» [En línea]. Disponible: <http://www.alegsa.com.ar/Dic/front-end.php>. [Último acceso: 11 Mayo 2013].
- [50] R. M. V. M. Leticia Betarte, PGMÚSICA Sistema de recomendación de Música, Universidad de la República, 2006.
- [51] «Master Thesis Seguido,» [En línea]. Disponible: <http://upcommons.upc.edu/pfc/bitstream/2099.1/7193/1/Master%20Thesis%20Seguido.pdf>. [Último acceso: 10 Abril 2013].
- [52] A. A. Ibáñez., *Sistema de recomendación de restaurantes georreferenciados*, España: Universidad de Jaén, Escuela Politécnica Superior de Jaén, Jaén.
- [53] T. P. O. A. G., *METODOLOGÍA DESARROLLO DE SOFTWARE PARA PYMES DE RETAI*, 2009.
- [54] M. C. H. Mejia, «La cultura orientada al cliente trasciende a una estructura de CRM en las instituciones de educación superior públicas y privadas (universidades) del eje cafetero,» *Universidad Nacional de Colombia*, 2008.

ANEXOS

ANÁLISIS E INTERPRETACIÓN DE LOS DATOS DE LA ENCUESTA REALIZADA

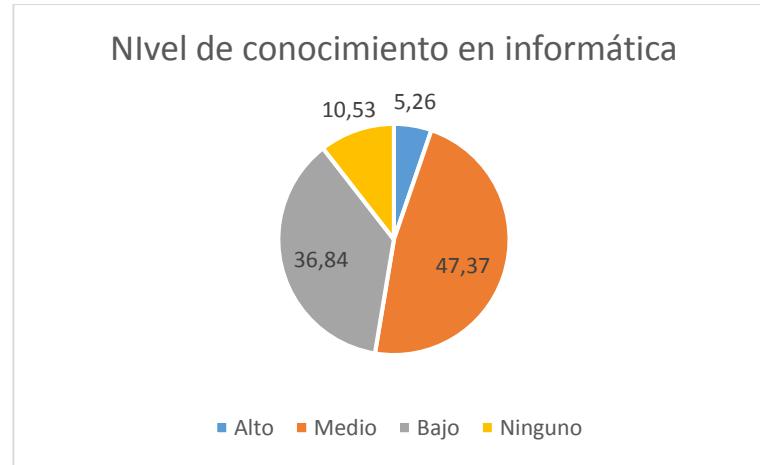
Para el análisis e interpretación del cuestionario aplicado, se usaron técnicas de procesamiento estadístico elemental, calculando frecuencias de respuestas y sus porcentajes.

El cuestionario del atipo cerrado se realizó en un total de 19 establecimientos comerciales de la ciudad de Bogotá Colombia, el proceso se realizó leyendo cada pregunta al usuario y anotando su respuesta.

A continuación se presentan los resultados obtenidos para cada pregunta, de igual forma se realiza la interpretación correspondiente:

- A la pregunta ¿Cuál es su nivel de conocimiento en informática? los resultados obtenidos fueron los siguientes:

Respuesta	Resultados	Porcentajes
Alto	1	5,26
Medio	9	47,37
Bajo	7	36,84
Ninguno	2	10,53



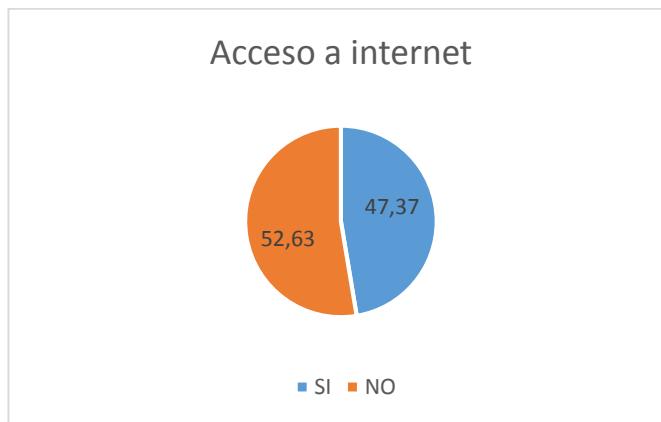
Interpretación:

Es claro que casi la mitad de los encuestados tiene conocimientos medios (47,37%) en informática lo suficiente para realizar tareas cotidianas en el comercio

electrónico, precedido de un 36,84% que tiene conocimientos bajos, conocimientos propios para tareas básicas en informática.

- A la pregunta ¿El establecimiento cuenta con acceso a internet? Tenemos:

Respuestas	Resultados	%
SI	9	47,37
NO	10	52,63



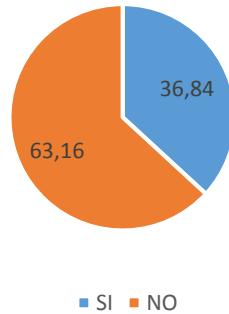
Interpretación:

Los resultados obtenidos tienen valores muy similares, casi la mitad de los establecimientos (52,63%) cuenta con acceso a internet de la misma manera mientras que el 47,37% no cuenta.

- A la pregunta ¿El establecimiento cuenta con correo electrónico? Tenemos:

Respuestas	Resultados	%
SI	7	36,84
NO	12	63,16

¿El establecimiento tiene E-mail?



Interpretación:

La mayoría (63,16%) de los establecimientos cuenta con un correo electrónico, que no necesariamente son de uso exclusivo para fines del comercio sino que también son de uso personal.

Contando con las respuestas que fueron afirmativas, se preguntó el correo electrónico con el fin de conocer el tipo de servicio que usaba, los resultados fueron los siguientes:

Respuesta	Resultados	%
Servicios gratuitos	6	85,71
pagos	1	14,29

Servicios gratuitos vs pagos

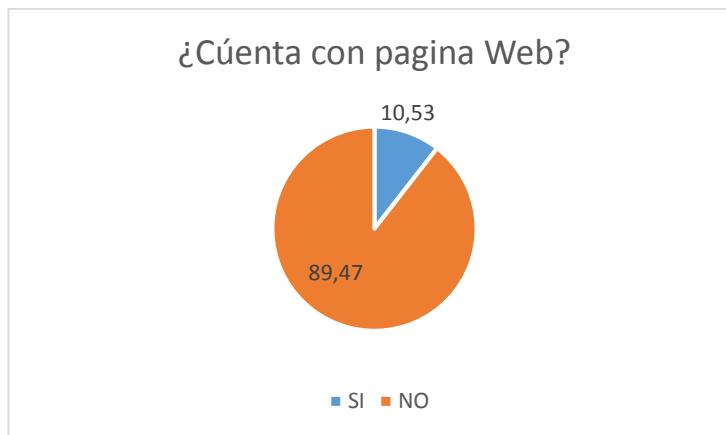


Interpretación:

Es claro que gran parte de los establecimientos que cuenta con correo suelen usar servicios gratuitos y conocidos (Yahoo, Gmail, Hotmail,..etc) a diferencia de correos con dominios propios que son pagos.

- A la pregunta ¿El establecimiento cuenta con página Web? Tenemos:

Respuestas	Resultados	%
SI	2	10,53
NO	17	89,47



Interpretación:

Claro está que es una minoría (10,53%) aquellos establecimientos que cuentan con una página web, casi el 90% de los encuestados no cuentan con una.

- A la pregunta ¿El establecimiento ha sido publicitado en algún medio? Tenemos:

Respuestas	Resultados	%
Clasificados	0	0,00
Páginas amarillas	1	5,26
Páginas amarillas internet	0	0,00
Otro	2	10,53
ninguno	16	84,21



Interpretación:

Es claro que la mayoría (84,21%) nunca a promocionado su establecimientos por algún medio, por otro lado otros medios (Pancartas, repartir volantes, perifoneo y megafoneo) son la principal vía de publicidad usado por los establecimientos, dejando atrás Internet.

- **A la pregunta ¿Considera usted importante que su establecimiento sea publicitado en internet de manera gratuita? Tenemos:**

Respuestas	Resultados	%
SI	14	73,68
NO	5	26,32



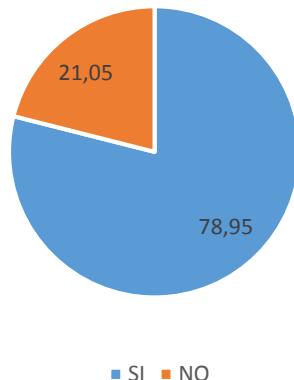
Interpretación:

El 73,68% de los encuestados convergen en que es importante publicitar su establecimiento en internet, viéndolo casi como una obligación de mercado.

- A la pregunta ¿Considera usted importa que información pública de su establecimiento (Nombre, dirección, teléfonos, actividad económica) pueda ser conocida a través de internet, para encontrar futuros clientes o nuevos negocios? Tenemos:

Respuestas	Resultados	%
SI	15	78,95
NO	4	21,05

¿Considera importante publicar información para conocer nuevos clientes?



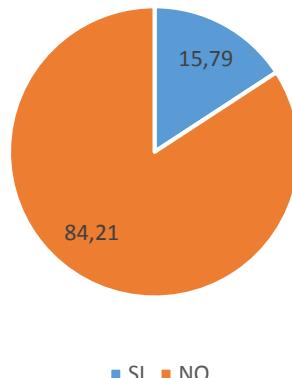
Interpretación:

La mayoría de los encuestados coinciden en que compartir información de sus establecimientos será un buen camino para conocer nuevos clientes y generar mayores negocios.

- A la pregunta ¿Tiene conocimientos o ha escuchado de los sistemas de recomendación en el comercio electrónico? Tenemos:

Respuestas	Resultados	%
SI	3	15,79
NO	16	84,21

¿Conoce los sistemas de recomendación en el comercio electrónico?



Interpretación:

De los encuestados tan solo un 15,79% tiene conocimientos de los sistemas de recomendación aplicados al filtrado colaborativo, el resto (84,21%) tienen un desconocimiento de este.

Análisis y conclusiones

Como ya se vio en la interpretación de los datos obtenidos, la mayoría de usuarios tiene un conocimiento de informática bueno, por lo tanto tienen las capacidades de realizar tareas básicas informáticas (Ingresar a una página web, usar un aplicativo, etc.) lo que justifica la implementación de un portal web ya que este tendrá los conocimientos para usarlo. Por otro lado se vio que un poco más de la mitad de establecimientos pagan un servicio de Internet, para nuestro caso no se vería como un problema debido a que el acceso a este es cada vez más fácil y no se necesita de pagar un servicio mensual para acceder a este.

Por otra parte el 73,68% de los encuestados convergen y tienen presente la importancia de publicitar su establecimiento, en especial en internet, viéndolo casi como una obligación y requerimiento para conocer nuevos clientes y generar nuevos negocios.

A pesar de lo anterior se pudo constatar que los propietarios aún no se “atreven” a entrar en Internet como medio publicitario de manera directa, un claro ejemplo es el pequeño porcentaje (10,53%) de establecimientos que cuenta con una página web propia o que alguna vez han publicitado en internet, esto se puede dar por

desconocimiento de herramientas para el comercio electrónico, por los costos de los servicios e incluso por el desconocimiento de su uso.