



EDUCACIÓN

SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Acapulco

TECNOLÓGICO NACIONAL DE MÉXICO
INSTITUTO TECNOLÓGICO DE ACAPULCO

SISTEMA INTELIGENTE PARA LA DETECCIÓN AUTOMÁTICA DE
COMPETENCIA Y MONITOREO DE PRODUCTOS EN MERCADO
LIBRE (MLCOMPET)

TESIS PROFESIONAL

QUE PARA OBTENER EL TÍTULO DE:
MAESTRO EN SISTEMAS COMPUTACIONALES

PRESENTA:
ING. GERARDO ALBERTO LÓPEZ VEGA

DIRECTOR DE TESIS:
M.T.I. JUAN MIGUEL HERNÁNDEZ BRAVO

CODIRECTOR DE TESIS:
DR. JOSÉ ANTONIO MONTERO VALVERDE

TUTOR DE TESIS:
DR. EDUARDO DE LA CRUZ GÁMEZ

ACAPULCO, GRO., DICIEMBRE 2020.

El presente trabajo de tesis fue desarrollado en la *División de Estudios de Posgrado e Investigación* del *Instituto Tecnológico de Acapulco*, perteneciente al Programa Nacional de Posgrados de Calidad (PNPC-CONACYT).

Con domicilio para recibir y oír notificaciones en Av. Instituto Tecnológico de Acapulco s/n, Crucero del Cayaco, Acapulco, Guerrero, México. C.P. 39905.

Becario:	Gerardo Alberto López Vega
CVU:	928453.
Núm. de apoyo:	927935.
Grado:	Maestría



DESCARGO DE RESPONSABILIDAD INSTITUCIONAL

Quien suscribe declara que el presente documento titulado "SISTEMA INTELIGENTE PARA LA DETECCION AUTOMATICA DE COMPETENCIA Y MONITOREO DE PRODUCTOS EN MERCADO LIBRE (MLCOMPET)" es un trabajo propio y original, el cual no ha sido utilizado anteriormente en instituci3n alguna para prop3sitos de evaluaci3n, publicaci3n y/o obtenci3n de alg3n grado acad3mico.

Adem3s, se adelanta que se han recogido todas las fuentes de informaci3n utilizadas, las cuales han sido citadas en la secci3n de referencias bibliogr3fica de este trabajo.

Acapulco, Gro; a 12 de diciembre de 2020.

GERARDO ALBERTO LOPEZ VEGA



Nombre y Firma

CARTA DE CESIÓN DE DERECHOS DE AUTOR

El que suscribe: GERARDO ALBERTO LOPEZ VEGA, autor del trabajo escrito de evaluación profesional en la opción de Tesis Profesional de Maestría con el título "SISTEMA INTELIGENTE PARA LA DETECCIÓN AUTOMÁTICA DE COMPETENCIA Y MONITOREO DE PRODUCTOS EN MERCADO LIBRE (MLCOMPET)", por medio de la presente con fundamento en lo dispuesto en los artículos 5, 18, 24, 25, 27, 30, 32 y 148 de la Ley Federal de Derechos de Autor, así como los numerales 2.15.5 de los lineamientos para la Operación de los Estudios de Posgrado; manifiesto mi autoría y originalidad de la obra mencionada que se presentó en la División de Estudios de Posgrado e Investigación, para ser evaluada con el fin de obtener el Título Profesional de Maestro en Sistemas Computacionales.

Así mismo expreso mi conformidad de ceder los derechos de reproducción, difusión y circulación de esta obra, en forma NO EXCLUSIVA, al Tecnológico Nacional de México campus Acapulco; se podrá realizar a nivel nacional e internacional, de manera parcial o total a través de cualquier medio de información que sea susceptible para ello, en una o varias ocasiones, así como en cualquier soporte documental, todo ello siempre y cuando sus fines sean académicos, humanísticos, tecnológicos, históricos, artísticos, sociales, científicos u otra manifestación de la cultura.

Entendiendo que dicha cesión no genera obligación alguna para el Tecnológico Nacional de México campus Acapulco y que podrá o no ejercer los derechos cedidos. Por lo que el autor da su consentimiento para la publicación de su trabajo escrito de evaluación profesional.

Se firma presente en la ciudad de Acapulco de Juárez, estado de Guerrero a los 07 días del mes de diciembre de 2020.

GERARDO ALBERTO LOPEZ VEGA



NOMBRE Y FIRMA



"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

Acapulco, Gro; a 7 de diciembre de 2020.

AUTORIZACIÓN DE IMPRESIÓN DE TESIS

Los abajo firmantes, miembros de la comisión revisora de tesis designada por la División de Estudios de Posgrado e Investigación del Tecnológico Nacional de México campus Acapulco para la evaluación de la tesis de la alumno GERARDO ALBERTO LOPEZ VEGA, manifiestan que después de haber revisado su tesis: "SISTEMA INTELIGENTE PARA LA DETECCIÓN AUTOMÁTICA DE COMPETENCIA Y MONITOREO DE PRODUCTOS EN MERCADO LIBRE (MLCOMPET)" desarrollada bajo la dirección del DIRECTOR, y el CO-DIRECTOR, el trabajo se **ACEPTA** para proceder a su impresión.

A T E N T A M E N T E


M.T.I. Juan Miguel Hernández Bravo
Nombre y firma del Director
Cédula Profesional: 6237953


Dr. José A. Montero Valverde
Nombre y firma del Co-director
Cédula Profesional: 5662755


DR. EDUARDO DE LA CRUZ GÁMEZ
Nombre y firma del Tutor
Cédula Profesional : 6526073



Enterado

Dr. Eduardo de la Cruz Gámez
Coordinador de la Maestría en Sistemas
Computacionales





Instituto Tecnológico de Acapulco
División de Estudios de Posgrado e Investigación

"2020, Año de Leona Vicario, Benemérita Madre de la Patria"

Acapulco Gro., 8/Diciembre/2020

NO. OFICIO: DEPI-207/2020

ASUNTO:
AUTORIZACIÓN DE
IMPRESIÓN DE TESIS PROFESIONAL

C. GERARDO ALBERTO LOPEZ

De acuerdo al reglamento de los Institutos Tecnológicos, dependiente de la Secretaría de Educación Pública y habiendo cumplido con todos los requisitos normativos respecto a su trabajo para titulación, Opción Titulación Tesis Profesional, con el proyecto titulado: SISTEMA INTELIGENTE PARA LA DETECCIÓN AUTOMÁTICA DE COMPETENCIA Y MONITOREO DE PRODUCTOS EN MERCADO LIBRE (MLCOMPET). Se **CONCEDE** la **AUTORIZACIÓN** para que proceda a la impresión del mismo.

Sin otro particular por el momento, me es grato quedar de usted.

A T E N T A M E N T E "
Educación Tecnológica con Competencias Social"



SECRETARÍA DE EDUCACIÓN
PÚBLICA
INSTITUTO TECNOLÓGICO
DE ACAPULCO
DIVISIÓN DE ESTUDIOS
DE POSGRADO E
INVESTIGACIÓN


EDUARDO DE LA CRUZ GÁMEZ
JEFE DE LA DIVISIÓN DE ESTUDIOS DE
POSGRADO E INVESTIGACIÓN

C.c.p. Expediente

EDG/stv



Tabla de Contenido

Índice de figuras	1
Índice de tablas	5
Índice de graficas	5
Índice de ecuaciones.....	5
Capítulo 1. Introducción	6
1.1 Antecedentes del Problema a Resolver.....	6
1.2 Planteamiento del Problema.....	7
1.3 Objetivos.....	9
1.3.1 Objetivo General	9
1.3.2 Objetivos Específicos	9
1.4 Justificación	10
1.5 Alcances y Limitaciones	11
1.6 Hipótesis	12
Capítulo 2. Antecedentes de la Investigación	13
2.1 Estado del Arte	13
2.2 Monitoreo Automatizado de Precios.....	13
2.3 Análisis de Datos de Transacciones.....	14
2.4 Detección y Minería Web.....	15
2.5 Data Warehouse y Toma de Decisiones	15
2.6 Aplicaciones de Software Relacionadas.....	16
2.6.1 Realtrends.....	17
2.6.2 Nubimetrics	18
2.6.3 EcomExperts.....	19
Capítulo 3. Marco Teórico	21
3.1 Análisis de competencia	21
3.2 La competencia en el comercio electrónico	24
3.3 Inteligencia artificial y los sistemas inteligentes.....	25
3.3.1 Características de los sistemas inteligentes.....	25
3.3.2 Tipos de sistemas inteligentes	27
3.3.3 Diferencia entre sistemas inteligentes y automatizados.....	27
3.4 Aprendizaje automático	29
3.4.1 Modelos.....	30
3.4.2 Clases de Algoritmos	30
3.4.3 Tipos de Algoritmos	31

3.5 Minería de datos	32
3.6 Descubrimiento de conocimiento en bases de datos (KDD).....	33
3.6.1 Etapas del KDD.....	34
3.7 Naive Bayes y el problema de la clasificación	35
3.9 Python y el análisis de datos	37
3.9.1 Django framework	37
3.9.2 Scikit Learn y Pandas	38
3.9.3 Naive Bayes con Scikit Learn.....	40
Capítulo 4. Metodología	41
4.1 Desarrollo del sistema	42
4.1.1 Diagrama de casos de uso.....	43
4.1.2 Diagrama de clase	44
4.2 Integración con Mercado Libre	46
4.2.1 Registro de la Aplicación.....	46
4.2.2 Autenticación.....	48
4.2.3 Autorización	50
4.2.4 Flujo de Conexión con Mercado Libre.....	52
4.2.5 Obtención de datos del usuario.....	54
4.2.6 Recurso de productos y búsquedas.....	55
4.2.7 Búsqueda de productos por Categoría	56
4.2.8 Búsqueda de productos por palabra clave.....	57
4.3 Preparación del dataset.....	59
4.3.1 Exploración de datos.....	60
4.3.2 Selección de datos.....	61
4.3.3 Limpieza de datos	61
4.3.4 Transformación de datos	61
4.3.5 Integración de datos.....	62
4.3.6 Formatear datos.....	62
4.4 Definición del Modelo	64
4.4.1 Validación del Modelo	67
4.5 Detección Automática.....	70
Capítulo 5. Desarrollo e Implementación	72
5.1 Desarrollo de la aplicación.....	72
5.1.1. Instalación de Python.....	73
5.1.2 Creación de Entorno Virtual para el proyecto.....	75
5.1.3 Instalación de Django.....	76

5.1.4	Instalación de otras librerías	79
5.1.5	Repositorio del proyecto en Github	79
5.1.6	Creación del proyecto con Django	82
5.1.7	Inicialización del proyecto	83
5.1.8	Levantar el servidor web del proyecto.....	84
5.1.9	Acceder al proyecto desde navegador web	85
5.1.10	Acceder al panel administrativo del proyecto.....	85
5.1.11	Configuración del Proyecto	87
5.2	Integración con Mercado Libre	88
5.2.1	Módulo de Mis Productos.....	89
5.2.2	Módulo Mis Búsquedas	93
5.3	Preparación del dataset.....	95
5.4	Definición del Modelo	98
5.4.1	Validación del Modelo	101
5.5	Detección automática	103
5.6	Preparación del ambiente de producción	105
5.6.1	Creación del servidor virtual privado (VPS).....	105
5.6.2	Descargar código del repositorio Github	108
5.6.3	Instalación de entorno virtual y dependencias	110
5.6.4	Configuración de Nginx y Unicorn.....	112
5.6.5	Modificar archivo settings del proyecto	114
5.6.6	Inicializar aplicación y base de datos	116
5.6.7	Configurar cuenta del usuario	118
5.6.8	Conectar aplicación a Mercado Libre	119
5.6.9	Obtener catálogo de productos del usuario.....	121
Capítulo 6	Pruebas y Resultados	123
6.1	Modulo de búsquedas	123
6.2	Modulo de detección de competencia	125
6.3	Modulo de seguimiento de productos	126
6.4	Modulo de actualización de precios.....	128
6.5	Análisis de Resultados	130
Conclusiones y trabajos futuros	132
Bibliografía	134

Índice de figuras

Figura 1 Real Trends.....	17
Figura 2 Nubimetrics	19
Figura 3 EcomExperts.....	19
Figura 4 Cuadro comparativo de competidores	23
Figura 5 Proceso de Obtención de Conocimiento (KDD).....	33
Figura 6 Librería pandas	39
Figura 7 Aplicación del algoritmo Naive Bayes con Scikit Learn.....	40
Figura 8 Metodología de Desarrollo	41
Figura 9 Plataforma de Comercio Electrónico Mercado Libre	42
Figura 10 Django Framework.....	43
Figura 11 Diagrama de Casos de Uso	44
Figura 12 Diagrama de clases	45
Figura 13 Proceso de Integración con Mercado Libre.....	46
Figura 14 Plataforma de Desarrollo de Mercado Libre.....	46
Figura 15 Administrador de Aplicaciones de Mercado Libre	47
Figura 16 Formulario de Información Básica de la Aplicación.....	47
Figura 17 Formulario de Alcance y Permisos de la Aplicación.....	48
Figura 18 Administrador de Aplicaciones de Mercado Libre	48
Figura 19 Autenticación del usuario en Mercado Libre	49
Figura 20 Autorización de acceso a Mercado Libre	50
Figura 21 Parámetro Code agregado a la URL de Autorización	50
Figura 22 Ejemplo de Access Token obtenido de Mercado Libre	52
Figura 23 Flujo de Conexión a Mercado Libre	53
Figura 24 Métodos disponibles para el recurso de Items y Búsquedas	55
Figura 25 Consulta de Productos por Categoría	57
Figura 26 Respuesta de llamada a API REST de productos por palabra clave	58
Figura 27 Preparación del Dataset.....	59
Figura 28 Python para Windows	73
Figura 29 Descarga de Python.....	74
Figura 30 Instalación de Python.....	74
Figura 31 Agregar Python al PATH.....	74
Figura 32 Verificar instalación de Python	75
Figura 33 Creación del entorno virtual para el proyecto.....	75

Figura 34 Estructura de archivos de un entorno virtual de Python	76
Figura 35 Activar entorno virtual de Python	76
Figura 36 Entorno virtual activo.....	76
Figura 37 Instalar Django con el gestor de paquete pip.....	77
Figura 38 Descarga e Instalación de Django	77
Figura 39 Librerías instaladas ejecutando el comando pip freeze	78
Figura 40 Generación de archivo de dependencias.....	78
Figura 41 Ubicación del archivo de dependencias.....	78
Figura 42 Github plataforma para control de versiones.....	80
Figura 43 Formulario de acceso a Github	80
Figura 44 Repositorio de Github para el proyecto.....	81
Figura 45 Controles del repositorio	81
Figura 46 Creación del proyecto con Django.	82
Figura 47 Proyecto creado con Django.....	82
Figura 48 Archivos de configuración del sistema	83
Figura 49 Inicialización del proyecto con el comando mígrate	83
Figura 50 Levantar servidor web del proyecto	84
Figura 51 Levantar servidor web en puerto 80.....	84
Figura 52 Aplicación corriendo desde el navegador web.....	85
Figura 53 Creación de super usuario del sistema.....	86
Figura 54 Formulario de Acceso al Admin del sistema	86
Figura 55 Panel de control del Admin del sistema	87
Figura 56 Configuraciones del proyecto.....	87
Figura 57 Archivo de configuración del proyecto	88
Figura 58 Librerías disponibles para integrar con Mercado Libre	88
Figura 59 Repositorio de Código en GitHub de Mercado Libre.....	89
Figura 60 Campo de Búsqueda Modulo de Mis Productos	90
Figura 61 Método para consultar productos por vendedor.....	90
Figura 62 Método de búsqueda de productos mayor a 1000 registros.	90
Figura 63 Ejemplo de consulta al API REST para el método de búsqueda	91
Figura 64 Datos obtenidos de Mercado Libre en formato JSON.....	91
Figura 65 Ejemplo de procesamiento de la respuesta obtenida y almacenada	91
Figura 66 Datos del producto vistos con panel administrativo de Django.	92
Figura 67 Tabla de Productos con Paginación.....	93
Figura 68 Módulo de Mis Búsquedas.....	93
Figura 69 Resultado de búsqueda del usuario.....	94

Figura 70 Vista general de los datos de productos	94
Figura 71 Datos de registros almacenados en la base de datos sin transformar.....	96
Figura 72 Datos de reputación del vendedor en formato JSON.....	97
Figura 73 Librerías de Python Scikit Learn y Pandas.....	99
Figura 74 Importación de librerías y módulos Scikit Learn y Pandas.....	99
Figura 75 Definición del Dataset con Pandas	99
Figura 76 Definición de Atributos y Clase en el Dataset	100
Figura 77 Asignación de Datos de Entrenamiento y Pruebas.....	100
Figura 78 Entrenamiento del Modelo de Clasificación	101
Figura 79 Predicción de datos de prueba.....	101
Figura 80 Obtención de métricas de desempeño del modelo	101
Figura 81 Matriz de confusión para la clase Competencia.....	102
Figura 82 Score de precisión del modelo de clasificación.....	103
Figura 83 Reporte de seguimiento de competencia y precio sugerido.....	104
Figura 84 Proveedor de servicios VPS Digital Ocean	105
Figura 85 Planes de VPS ofrecidos por Digital Ocean.....	106
Figura 86 Diferentes sistemas operativos ofrecidos por Digital Ocean	106
Figura 87 Plan seleccionado para poner el proyecto en producción.....	107
Figura 88 Servidor VPS configurado en Digital Ocean.....	108
Figura 89 Conexión SSH al servidor del proyecto.....	108
Figura 90 Clonar repositorio del proyecto en Github.....	108
Figura 91 URL repositorio del proyecto en Github	109
Figura 92 Código fuente descargado en el servidor.....	109
Figura 93 Entorno virtual creado en servidor de producción.....	110
Figura 94 Instalación de dependencias del proyecto	111
Figura 95 Instalación de dependencias del proyecto.	111
Figura 96 Iniciar servidor web del proyecto con Gunicorn.....	112
Figura 97 Archivo WSGI del proyecto	113
Figura 98 Instalación de Nginx.....	113
Figura 99 Archivo de configuración de Nginx para el proyecto.	114
Figura 100 Configuración de host permitidos al proyecto en producción.....	114
Figura 101 Configuración de constantes utilizadas en el código.....	115
Figura 102 Conexión a la base de datos sqlite3	116
Figura 103 Inicializar base de datos en producción	116
Figura 104 Creación del super usuario en producción.....	116
Figura 105 Acceso al panel administrativo en producción	117

Figura 106 Base de datos inicializada en producción.	117
Figura 107 Apartado de autenticación y autorización del proyecto	118
Figura 108 Creación de cuenta de usuario.	118
Figura 109 Menú Perfil de la aplicación	119
Figura 110 Conexión a Mercado Libre	120
Figura 111 Autorización de acceso a la aplicación a la cuenta de Mercado Libre .	120
Figura 112 Cuenta vinculada a Mercado Libre.....	121
Figura 113 Búsqueda de productos en el catálogo del usuario.....	121
Figura 114 Catálogo de productos del usuario.....	122
Figura 115 Módulo de Mis Búsquedas	123
Figura 116 Resultado de búsqueda del usuario.....	124
Figura 117 Vista general de los datos de productos	124
Figura 118 Listado de resultados de búsqueda	125
Figura 119 Detección de competencia	126
Figura 120 Seguimiento de los productos de la competencia	126
Figura 121 Listado de productos en seguimiento.....	127
Figura 122 Detalle del seguimiento de la competencia	127
Figura 123 Sugerencias y Cambio de Precio en productos	128
Figura 124 Actualización de precios en los productos seleccionados.....	129

Índice de tablas

Tabla 1 Métodos disponibles para la librería GaussianNB.....	40
Tabla 2 Listado de dependencias	73
Tabla 3 Dependencias del proyecto	79
Tabla 4 Datos de entrenamiento para el modelo de clasificación	98

Índice de graficas

Grafica 1 Resultados de tiempo invertido en la detección de competencia	130
Grafica 2 Resultados de numero de productos monitoreados	131
Grafica 3 Resultados del tiempo de respuesta para actualizar precios.....	131

Índice de ecuaciones

Ecuación 1 Teorema de Bayes	65
Ecuación 2 Ejemplo de probabilidad compuesta.....	65
Ecuación 3 Probabilidad compuesta aplicando probabilidad condicional.....	65
Ecuación 4 Aplicación de ingenuo de Naive Bayes	66
Ecuación 5 Probabilidad compuesta	66
Ecuación 6 Naive Bayes expresado con distribución condicional	66

Capítulo 1. Introducción

1.1 Antecedentes del Problema a Resolver

Actualmente vivimos en un mundo demasiado competitivo, en donde las empresas lanzan al mercado nuevos productos cada vez con mayor frecuencia, y aparecen cada día nuevas empresas competidoras, las cuales con ayuda de la tecnología han podido desarrollar estrategias y modelos de negocio que les permiten diferenciarse de su competencia y ganar terreno en el mercado.

En la actualidad, el comercio electrónico ha cambiado por completo la forma tradicional de hacer negocios y ha influido tanto en vendedores como clientes. Este significativo cambio en el modelo de negocio ha experimentado un tremendo crecimiento en todo el mundo gracias a la penetración masiva en internet ya que ha contribuido al crecimiento las nuevas empresas que han estado utilizando esta opción como un modelo de negocio diferenciador (Dr. Shahid & Kansana, 2016).

De acuerdo con la Encuesta Nacional sobre Disponibilidad y Uso de Tecnologías de la Información en los Hogares (ENDUTIH) 2018, en los últimos años ha ido al alza el interés de los mexicanos por el comercio electrónico, y actualmente en el país se calcula que cerca de 20 por ciento de las personas que tienen acceso a Internet lo usa para ordenar o comparar productos (ENDUTIH, 2018).

Una de las principales plataformas de comercio electrónico en México es Mercado Libre, la cual alcanza los 20 millones de visitas mensuales. Según datos de Mercado Libre el aumento de vendedores en la plataforma ha tenido un

comportamiento exponencial en los últimos años llegando a alcanzar 10.8 millones de vendedores activos actualmente (MercadoLibre, 2018).

El creciente aumento de competidores en la plataforma representa un reto importante para empresas que utilizan este medio para realizar sus transacciones sobre todo para las pequeñas y medianas empresas (PyMES), ya que cuentan con recursos limitados tanto humanos como financieros por lo que, para competir con empresas más grandes y posicionadas, es necesario que realicen un adecuado análisis de su competencia, así como también que cuenten con las herramientas que les permitan optimizar sus recursos y ser más eficientes en sus procesos.

1.2 Planteamiento del Problema

PROMUSICGEAR es una empresa distribuidora de instrumentos musicales y accesorios la cual utiliza distintos canales para vender sus productos, entre ellos su tienda física, tienda en línea y plataformas de comercio electrónico como Mercado Libre, siendo esta última la que mayor volumen de transacciones realiza.

La problemática presentada por la empresa es que actualmente sus ingresos se han visto afectados ya que el volumen de ventas ha disminuido, esto debido al aumento de los competidores en la plataforma que ofrecen los mismos productos con precios similares y en algunos casos más competitivos.

Debido a esta situación, la empresa se ha visto en la necesidad de hacer un continuo análisis y seguimiento de su competencia, monitoreando de manera constante los productos que ofrecen sus competidores de modo que puedan reaccionar y ofrecer una mejor alternativa para sus clientes.

Para llevar a cabo este proceso, actualmente cada colaborador debe de encargarse de darle seguimiento a ciertos productos y competidores. Posteriormente, para cada uno de los productos que sigue, se debe de hacer un análisis de diferentes factores como el precio de venta, costo, stock y margen de ganancia para poder así tomar decisiones y de ser necesario replantear una mejor estrategia de ventas. Sin embargo, llevar a cabo este proceso tiene un alto costo humano y económico, ya que debe de invertir bastante tiempo en realizar esta tarea, razón por la cual llegan a descuidar otras actividades importantes.

Por otro lado, debido al alto número de productos que manejan no es posible darles un seguimiento eficiente a todos ellos por lo que deben seleccionar sólo algunos de los productos que consideran más relevantes y que representan un mayor impacto en sus ventas.

Es por eso que al contar con una herramienta que les permita detectar automáticamente su competencia y monitorear los productos, les permitiría tomar mejores decisiones y desarrollar estrategias de manera más eficiente, con las cuales podrá diferenciarse y competir contra las demás empresas.

1.3 Objetivos

1.3.1 Objetivo General

Desarrollar un sistema inteligente que permita detectar competidores de manera automatizada dentro de la plataforma de ventas Mercado Libre, el cual permita al vendedor monitorear de manera eficiente los productos publicados y así, reaccionar rápidamente para establecer nuevas estrategias de venta que le permitan mejorar el posicionamiento de sus productos frente a los de sus competidores.

1.3.2 Objetivos Específicos

- Desarrollar una aplicación de software con el que se comunique con la plataforma Mercado Libre a través del consumo de su API REST.
- Desarrollar un módulo que permita hacer búsquedas de los distintos productos dentro de la plataforma Mercado Libre y listar la información obtenida.
- Desarrollar un módulo que permita hacer una selección de los productos listados y registrarlos para monitorear sus precios.
- Desarrollar un módulo que aplique un algoritmo de clasificación con el que se pueda detectar automáticamente un producto de la competencia.
- Desarrollar una funcionalidad donde se establezcan criterios bajo los cuales se puedan actualizar precios de manera automática.
- Desarrollar una funcionalidad que se ejecute periódicamente la cual debe de comparar los precios entre los diferentes productos para determinar cambios.
- Desarrollar un módulo de reportes que indique que productos han sufrido cambios y haga sugerencias sobre la actualización de precios.

1.4 Justificación

Las PyMES cumplen un importante papel en la economía de todos los países. Actualmente en México existen 4 millones de PyMES activas, las cuales constituyen el 80% de las empresas, el 79% del empleo, y contribuyen con un 52% del PIB (Forbes, 2018).

Según la empresa Linio, en su informe sobre el Índice Mundial de Comercio Electrónico, las ventas de comercio electrónico en América Latina superan los 57,000 millones de dólares (Linio, 2018). De acuerdo con la consultora ComScore, el comercio electrónico en América Latina se encuentra dominado por Mercado Libre, con 56.3 millones de visitantes al mes (ComScore & EMarketer, 2018).

De acuerdo a Forbes, uno de los principales retos para las PyMES en México es la innovación tecnológica, solo el 6% de las PyMES aprovecha las TIC's para mejorar su productividad (Forbes, 2018).

Por lo anterior es de gran importancia contar con las herramientas tecnológicas necesarias que le brinden información veraz y oportuna del estatus actual de la organización con la cual puedan tomar decisiones y desarrollar estrategias que les permitan crecer y tener un mejor aprovechamiento de sus recursos. La presente investigación surge de la necesidad de mejorar el posicionamiento de los productos de la empresa en la plataforma de ventas Mercado Libre mediante el desarrollo de un software que le permita detectar competidores de manera automática y así poder reaccionar rápidamente a los cambios en el comportamiento de sus productos, además de servirle de apoyo para replantear nuevas estrategias comerciales que le permitan mejorar el posicionamiento de la empresa frente a sus competidores.

1.5 Alcances y Limitaciones

El desarrollo de este proyecto considera cubrir los objetivos mencionados anteriormente al desarrollar una aplicación de software que ayude a los vendedores de Mercado Libre a detectar automáticamente a su competencia, además de que les permita monitorear las publicaciones y proporcione alternativas para mejorar el precio de sus productos.

El sistema plantea la utilización de la técnica de minería de Datos Naive Bayes con la finalidad de desarrollar un modelo de clasificación, el cual, tras analizar los datos referentes a las variables de los productos como la reputación del vendedor, el precio del producto, el tipo de publicación, la forma de pago y el tipo de envío, permita determinar si un producto publicado es considerado competidor.

Los procesos dentro de una empresa son diversos dependiendo del tipo de organización que se trate, hacer uso de diferentes algoritmos de inteligencia artificial y técnicas para la minería de datos, abren un abanico amplio de posibilidades para que puedan desarrollarse más herramientas que ayuden a automatizar y hacer más eficientes los procesos dentro de las empresas. La utilización de otros modelos y técnicas de minería de datos quedan fuera del alcance de este trabajo y quedan abiertas para su utilización en proyectos futuros.

1.6 Hipótesis

Al implementar un sistema inteligente para la detección automática de competencia y monitoreo de productos en Mercado Libre se podrá reducir el tiempo invertido para la detección y seguimiento de la competencia hasta en un 50%, así como también permitirá hacer un seguimiento a una mayor cantidad de productos.

Capítulo 2. Antecedentes de la Investigación

2.1 Estado del Arte

La industria actual está conformada por el uso de diversas tecnologías, las cuales han originado una gran transformación en la forma de hacer los negocios. La generación de grandes cantidades de datos diariamente gestionados con Big Data, las numerosas aplicaciones de software gestionadas con Cloud Computing y la utilización de la inteligencia artificial en cada vez más proyectos han dado lugar a un desarrollo que hasta ahora solo habíamos podido imaginar.

De acuerdo con un estudio de la consultora McKinsey sobre automatización se determinó que hasta un 40% del tiempo que se dedica a las actividades relacionadas con las ventas se podrían automatizar al hacer uso de la inteligencia artificial (IA). El conocer las preferencias de los clientes, procesar transacciones, tomar pedidos, pronosticar ventas y personalizar las ofertas por cliente o sector son algunos ejemplos de que actividades pueden mejorarse gracias al uso de la IA (McKinsey & Company, 2018).

2.2 Monitoreo Automatizado de Precios

Un ejemplo de aplicación de la IA en las ventas es el monitoreo de precios de manera automatizada, esto con el fin de permitir conocer la posición de la empresa frente a sus competidores (Radoslav Fasuga, 2014). Este artículo trata el problema de la determinación automatizada del precio de venta del producto en el entorno en línea, también describe el método de cálculo y la importancia de la posición del

mercado global en función de los precios y los servicios ofrecidos, además describe las fuentes donde puede obtener información sobre los precios y la comercialización de los productos. Describe el procedimiento para encontrar los productos relevantes en los sitios web de la competencia, en los agregadores de productos y en los sistemas de subastas.

El artículo describe el monitoreo de tendencias de precios, disponibilidad de productos, ofertas promocionales, ventas y otros medios de promoción de ventas y marketing en línea. En su parte final, el artículo describe el método que, basándose en la información obtenida, recomienda establecer el nivel de precio del producto en particular, monitorea la evolución del mercado y recomienda los cambios en el precio del vendedor en relación con el estado actual del mercado.

2.3 Análisis de Datos de Transacciones

Otra aplicación de la IA en las ventas es en el análisis y la optimización de ventas de productos online (Pirani, 2017). Este artículo habla de que la enorme cantidad de sitios web de comercio electrónico en el mercado actual ha llevado a la necesidad de contar con herramientas de análisis que ayuden a determinar si la organización cumple con los objetivos de ventas deseados. El documento plantea desarrollar una "Herramienta de análisis de ventas" y describe la necesidad de un sistema para analizar las transacciones de la base de datos de los sitios web de comercio electrónico utilizando diversas técnicas y algoritmos de extracción de datos, como el análisis de afinidad, la regresión logística y la regresión lineal.

La propuesta anterior plantea analizar los datos de los productos vendidos, posteriormente segmentar los datos obtenidos y analizar los resultados para extraer las tendencias del mercado y los patrones de ventas de productos. Al finalizar, el

sistema optimiza estos datos sobre la base de los requisitos del mercado, lo que mejora la planificación de ventas y mercaderías de manera que aumente la productividad general y los beneficios de la organización.

2.4 Detección y Minería Web

Otra forma para mejorar de las ventas es mediante la detección y el almacenamiento de objetos (Sharma Mayank, 2016). En este proyecto detectar el logotipo del producto y recopilar su información de los archivos XML ubicados en la base de datos remota mediante el uso de Haar Cascades. Adicionalmente se emplea la técnica del web-scraping para obtener los datos almacenados en el archivo XML y se muestra la información que puede incluir las revisiones, el precio, el contacto o el sitio web oficial para realizar el pedido. En esta aplicación se utiliza el aprendizaje automático ADA-BOOST para pesar las imágenes positivas y negativas y separarlas de acuerdo con los requisitos del logotipo. Además, mediante el uso del algoritmo de Cam-Shift se puede hacer un seguimiento en tiempo real.

Esto ayuda a aumentar las ventas donde el usuario que olvida el producto a comprar si no tiene información previa sobre el producto. El uso del almacenamiento de objetos nos permite acceder a una gran cantidad de datos del producto a través de la API REST en el tiempo más rápido, lo que reduce la complejidad del tiempo y la latencia de la recopilación de información.

2.5 Data Warehouse y Toma de Decisiones

Otra aplicación importante es en el análisis y toma de decisiones basado en Data Warehouse (Changhui, 2016). El documento plantea que el análisis de los sistemas de tecnología de almacenamiento de datos, el procesamiento y

comunicación de estos datos son un desafío, sobre todo los sitios web de compras en línea que se inundan con una gran cantidad de datos de ventas todos los días, por lo que poder analizarlos y visualizarlos es una tarea de mucha importancia.

Actualmente, los datos se generan muy rápidamente debido al aumento de la información en la vida cotidiana. Se acumula una gran cantidad de datos de varias organizaciones que es difícil de analizar y explotar. Los datos creados por un número creciente de sensores en el entorno, como cámaras de tráfico y satélites, actividades de Internet en sitios de redes sociales, bases de datos de atención médica, bases de datos gubernamentales, datos de ventas, etc., son ejemplos de datos enormes.

El Data Warehouse es un entorno de datos estructurado del sistema de soporte de decisiones y la fuente de datos de procesamiento analítico en línea. En este documento se describen las características y el método de diseño del modelo de datos en el Data Warehouse, posteriormente se introduce un sistema de soporte de decisiones basado en el Data Warehouse, tomando el establecimiento del sistema de análisis de decisiones de ventas, por ejemplo, el análisis combinado con OLAP, proporciona un diseño e implementación de métodos basados en decisiones.

2.6 Aplicaciones de Software Relacionadas

Mercado Libre con el fin de mejorar la experiencia y de brindarles más herramientas a sus usuarios cuenta con una tienda en línea de aplicaciones las cuales están orientadas a diferentes enfoques, por ejemplo. Publicar y sincronizar masivamente, comunicación con clientes, administración del negocio, gestión de ventas, análisis de mercados y publicidad de artículos.

A continuación, se analizan algunas de las aplicaciones más relevantes que se han considerado como referencia para el desarrollo del presente proyecto.

2.6.1 Realtrends

Una de las aplicaciones más populares entre los usuarios se llama REALTRENDS la cual cuenta con más de 1000 descargas y usuarios que la utilizan. Esta aplicación se especializa en las siguientes áreas, por ejemplo, publicar y sincronizar masivamente, comunicación con clientes, gestión de ventas y análisis de mercados (Figura 1).

Una de las características interesantes de esta aplicación es la parte de análisis de mercado y seguimiento de la competencia, donde permite hacer un seguimiento a los usuarios que sean sus competidores, seleccionándolos y monitoreando sus movimientos, ventas y actualización de precios con el fin de analizar su comportamiento y el usuario pueda reaccionar rápidamente para aplicar alguna estrategia que permita estar mejor posicionado que ellos.



Figura 1 Real Trends

Esta herramienta tiene un costo por volumen de ventas, es decir entre más ventas tenga el costo es mayor. El costo puede ser mensual o anual y van desde los \$279.00 pesos al mes cuando son 15 ventas o menos y hasta \$1519.00 pesos por

1,000 o más ventas, esto para el costo mensual. El costo anual maneja un descuento de 20% sobre los costos mencionados.

Aunque esta herramienta es buena y de mucha utilidad en algunos casos, durante la práctica resulta difícil gestionar de manera efectiva toda la información que genera para así reaccionar de manera rápida conforme se vayan presentando los cambios en los productos de los competidores, esto se complica más cuando el número de productos que se venden por esta plataforma es elevado, tal cual es el caso de estudio de este proyecto, en donde la empresa cuenta con más de 5000 productos y el proceso de identificación y selección de competidores, monitoreo de productos y análisis de la información obtenida se vuelve una tarea en donde la complejidad y el tiempo aumentan de forma exponencial, por lo tanto la efectividad en el tiempo de reacción se va mermando conforme se van agregando más datos a la aplicación. Por esa razón esta herramienta no logra cumplir con el objetivo principal que es aumentar la competitividad de la empresa.

2.6.2 Nubimetrics

Nubimetrics es otra aplicación con funcionalidades que le permiten a los usuarios conocer información importante acerca de sus competidores, lo cual es de gran ayuda para que conozcan la posición de su empresa, así como conocer las estrategias que están aplicando. Esta aplicación pretende que los vendedores conozcan a sus competidores para poder adelantarse a sus estrategias, para lo cual les permite compararse de manera simultánea con ellos. Con esta aplicación se puede analizar ventas en dinero, unidades vendidas, visitas y conversión para entender en dónde los vendedores deben mejorar (Figura 2).



Figura 2 Nubimetrics

Una de las características importantes que cabe resaltar de esta aplicación es el tema de compararse simultáneamente con múltiples competidores y la diversidad de atributos que compara como el número de ventas, lo cual permite valorar una publicación no solo por su precio si no por otro tipo de atributos.

Los costos que maneja esta aplicación son en base a una entrevista mediante la cual se hace un cálculo del uso y características que se vayan a utilizar, por lo que no están publicados en su sitio web.

2.6.3 EcomExperts

EcomExperts es una plataforma que ofrece una alternativa bastante interesante a sus usuarios con el fin de mejorar sus ventas ya que una de sus principales características es que le permite conectarse a distintas aplicaciones mediante las cuales el vendedor puede gestionar todos sus canales de ventas. Por otro lado, esta plataforma está pensada para vendedores online ya que le permite también gestionar su almacén y automatizar la sincronización con todos sus canales de ventas como Mercado Libre y Linio.



Figura 3 EcomExperts

Algunas de las características que ofrece esta aplicación son la gestión de ventas, y compras, gestión de inventario y productos, gestión pagos y envíos, integraciones con Mercado Libre, Linio, Tienda Nube, entre otras. Los costos que maneja esta aplicación van en función del volumen de ventas mensual y van desde \$269.00 pesos al mes con un volumen de hasta 100 transacciones y \$699.00 pesos por un volumen mayor a 1,000 transacciones.

Uno de los puntos más fuertes de esta aplicación es la posibilidad de conectarse con múltiples aplicaciones lo cual podría ser bastante útil ya que la empresa actualmente cuenta con diversos canales de venta que podrían integrarse en una misma aplicación y así mejorar la gestión del negocio teniendo una vista más general de la situación de la empresa.

Capítulo 3. Marco Teórico

Es necesario conocer algunos conceptos empleados en el análisis de la competencia, así como en el funcionamiento de los sistemas inteligentes, de tal forma que se puedan establecer las características que sirvan para realizar una correcta detección de la competencia a través del análisis de los datos obtenidos.

3.1 Análisis de competencia

De acuerdo con la comisión Federal de Competencia (COFESE), la definición de competencia significa rivalidad entre empresas que participan en un mercado aplicando sus mejores estrategias de manera que pueden minimizar sus costos, maximizar sus ganancias y así mantenerse activas e innovadoras frente a otras empresas rivales (COFECE, 2018).

La competencia puede ser directa o indirecta, la competencia directa son aquellas empresas que producen o venden un producto igual o similar al nuestro y que además los venden en el mismo mercado que nosotros, es decir, que buscan los mismos clientes para venderles el mismo producto. Por otro lado, la competencia indirecta la conforman todas aquellas empresas que intervienen de forma lateral al mercado y clientes de los que se encuentra la empresa, y que buscan satisfacer las mismas necesidades, pero de otra forma o con otros productos (Endeavor, 2010).

El análisis de la competencia es el análisis de los recursos, capacidades, estrategias, ventajas competitivas, fortalezas, debilidades y demás características de los actuales y potenciales competidores de una empresa, que se realiza con el fin de poder, en base a dicho análisis, tomar decisiones o formular estrategias que permitan

competir con ellos de la mejor manera posible. Realizar el análisis de la competencia permite a las empresas estar prevenidas ante las nuevas acciones o estrategias de sus competidores, además de aprovechar sus debilidades, con el fin de bloquear o hacer frente a sus fortalezas, y tomar como referencia sus productos o las estrategias que mejores resultados les estén dando (CreceNegocios, 2019).

Un ejemplo puede ser que al analizar la competencia se descubre que una de las debilidades del competidor es su incapacidad para mantener precios bajos o su mala atención al cliente, en respuesta, la empresa puede tomar la decisión de reducir sus precios y dar prioridad a la atención al cliente con el fin de ganarles mercado. En el caso de las PyMES, el análisis va dirigido solamente a su competencia directa.

Realizar un análisis de la competencia no requiere de mucha complejidad y en algunos casos puedes llegar a resultar demasiado sencillo, a continuación, se muestran las etapas para poder realizar el análisis de competencia:

1. Determinar la necesidad del análisis:

Este primer paso consiste en determinar el objetivo del análisis, es decir el porqué de su realización, el objetivo puede ser distinto dependiendo de si la empresa es de reciente creación o si se trata de un negocio que ya se encuentra en marcha.

2. Determinar la información se va a recolectar:

Esta etapa consiste en determinar en base a la etapa anterior que información se va a necesitar de los competidores, por ejemplo, número de competidores, su ubicación, sus volúmenes de ventas, sus productos, sus precios, recursos, etc.

3. Obtener la información:

Ya que se ha determinado la información que se va a recolectar, el siguiente paso consiste en realizar la recolección. Es muy común emplear para esta etapa la técnica de observación, por ejemplo, visitar sus establecimientos y tomar nota de

sus productos y precios. Otra forma muy utilizada para recolectar información de los competidores es mediante la aplicación de entrevistas o encuestas informales a sus actuales o antiguos trabajadores, proveedores o clientes. Por último, otra forma bastante efectiva y comúnmente utilizada actualmente es buscar en internet ya sea revisando la información de sus páginas web, revisar su información en redes sociales, leer los comentarios de sus clientes, entre otras.

4. Analizar la información obtenida:

Una vez que la información ha sido recolectada, el siguiente paso consiste en analizarla, para ello se recomienda elaborar un cuadro comparativo de los competidos y la valoración de los factores que se consideren mas importantes para el análisis como puede observarse a continuación (Figura 2).

	Principales fortalezas	Principales debilidades	Principales estrategias
Competidor A	Buena calidad del producto	Falta de capacidad para mantener precios bajos	Activa participación en redes sociales
Competidor B	Variedad de diseños	Mala atención al cliente	Precios bajos
Competidor C	Buena ubicación	Poca calidad de sus insumos	Especialización en un tipo de producto

Figura 4 Cuadro comparativo de competidores

5. Toma de decisiones y formulación de estrategias

Después de analizar la información obtenida se procede a tomar decisiones o formular estrategias que permitan a la empresa aprovechar las oportunidades o hacer frente a las amenazas encontradas en el análisis. Un ejemplo de decisión puede ser que se descubre que el competidor no puede mantener sus precios bajos por el alto costo de operación o fabricación del producto, se puede tomar la decisión de reducir sus precios con el fin de bloquearlos y ganarles mercado.

El análisis de la competencia es una tarea muy importante para las empresas y no debe realizarse solo una vez, sino que debe hacerse de manera constante. Adicionalmente al proceso descrito anteriormente, también hay que considerar estar siempre atentos a las acciones de los competidores, por ejemplo, el lanzamiento de un nuevo producto o la aplicación de nuevas estrategias publicitarias, entre otras, con el fin de adelantarse a dichas estrategias.

3.2 La competencia en el comercio electrónico

El comercio electrónico, también conocido como e-commerce (*electronic commerce* en inglés), consiste en la compra y venta de productos o servicios a través de internet, ya sea por redes sociales y páginas web. La penetración masiva en Internet ha contribuido al crecimiento del comercio electrónico y particularmente, las nuevas empresas han estado utilizando cada vez más esta opción como un modelo de negocio diferenciador (Dr. Shahid & Kansana, 2016).

Hoy en día el comercio electrónico ha crecido de manera extraordinaria debido a Internet, en 2017, dos mil millones de personas realizaron una transacción de comercio electrónico móvil (Toptal, s.f.), de acuerdo con la Comisión Nacional de los Mercados y la Competencia (CNMC) ha alcanzado una tasa de crecimiento interanual de un 23,4% (WikiSaber, s.f.).

El comercio electrónico cuenta con diferentes ventajas dependiendo el rol que juegue el usuario en la transacción. Para el caso de las empresas, el comercio electrónico realizado entre 2 empresas recibe el nombre de B2B (*business-to-business* en inglés). Algunas de las ventajas de este tipo de comercio son:

- Simple distribución
- Comunicación vía electrónica

- Conocer mejor a sus clientes
- Fidelizar clientes

En el caso de los usuarios, algunas de las principales ventajas son:

- Comparar productos y encontrar un producto a menor costo.
- Negociación con el vendedor.
- Comodidad en la adquisición del producto o servicio.

3.3 Inteligencia artificial y los sistemas inteligentes

Kamplan y Haenlein definen la inteligencia artificial como "*la capacidad de un sistema para interpretar correctamente datos externos, para aprender de dichos datos y emplear esos conocimientos para lograr tareas y metas concretas a través de la adaptación flexible*" (Kamplan & Haenlein, 2019).

Un sistema inteligente es un sistema de software que tiene algunas de las características similares a la inteligencia mostrada por las personas, estos pueden recibir información obtenida de su entorno y responder en base a la información almacenada previamente en su memoria. Los sistemas inteligentes pueden mejorar con base en la experiencia obtenida y aprender para arrojar resultados más precisos. Un ejemplo son los sistemas para detectar spam en los correos electrónicos, los cuales mejoran cada vez más con la información que se obtiene de los usuarios.

3.3.1 Características de los sistemas inteligentes

Como se mencionó anteriormente, para que un sistema puede ser considerado inteligente debe contar con algunas características:

- **Inteligencia**, se refiere a la capacidad para obtener información nueva utilizando tanto la información previamente registrada, así como la obtenida del exterior.
- **Sistematización**, se refiere a que un sistema inteligente trabaja como parte de un sistema, el cual está fuertemente relacionado a cada uno de los componentes que lo integran.
- **Objetivo**, se refiere a aquello para lo que el sistema fue desarrollado, es decir, lo que busca obtener con el sistema a partir de los datos que se le hayan proporcionado.
- **Capacidad sensorial**, es la capacidad para recibir y procesar información de su entorno para utilizarla y realizar una acción en base a los datos obtenidos.
- **Conceptualización**, se refiere a abstraer la información y darle un sentido para generar un concepto. Los conceptos generados están interrelacionados para darle sentido a la información.
- **Reglas de actuación**, se refiere a las acciones que ha de realizar el sistema en base a una experiencia o de lo que se haya interpretado en la memoria, en otras palabras, hace una relación entre la condición y la acción a realizar.
- **Memoria**, es el espacio de almacenamiento físico en donde se guardan los conceptos y las reglas, así como también la nueva experiencia que se vaya obteniendo.
- **Aprendizaje**, se considera la parte más importante de un sistema inteligente y es el resultado de la conjunción de las otras características, el aprendizaje se considera la capacidad para detectar un patrón en la situación o condición presentada y determinar la acción futura que ha de realizarse.

3.3.2 Tipos de sistemas inteligentes

De acuerdo con Stuart Russell y Peter Norvig a lo largo de la historia se han seguido cuatro enfoques los cuales se dividen entre los que están centrados en los humanos y aquellos centrados en la racionalidad (Peter & Russell, 2009)

Los que se enfocan en los humanos son:

- **Sistemas que piensan como humanos.** Estos sistemas tratan de emular el pensamiento humano; por ejemplo, las redes neuronales artificiales
- **Sistemas que actúan como humanos.** Estos sistemas tratan de actuar como humanos; es decir, imitan el comportamiento humano; por ejemplo, la robótica

Los que se enfocan en la racionalidad:

- **Sistemas que piensan racionalmente.** Es decir, con lógica (idealmente), tratan de imitar o emular el pensamiento lógico racional del ser humano; por ejemplo, los sistemas expertos.
- **Sistemas que actúan racionalmente (idealmente).** Tratan de emular de forma racional el comportamiento humano; por ejemplo, los agentes inteligentes.

3.3.3 Diferencia entre sistemas inteligentes y automatizados.

Un sistema automatizado hace exactamente aquello para lo que ha sido programado previamente y nada más. En este sentido, la automatización es ideal para tareas repetitivas y monótonas en las que, tanto la información que debe procesar el sistema como los diferentes pasos de la tarea a realizar son perfectamente previsibles y limitados.

Tareas que para un humano pueden resultar tediosas, en las que se puede acabar cansado por aburrimiento y cometer un error, o que requieren de movimientos repetitivos que podrían acabar causándole una lesión física, un sistema automatizado puede realizar millones de veces la misma operación o movimiento y, no solo no se cansa, sino que lo puede hacer con mayor precisión y rapidez que un operador humano. A diferencia de los sistemas automatizados, la inteligencia artificial puede identificar patrones en un conjunto de datos y optar por la que considera la mejor respuesta o solución en función de los datos de los que dispone, sin haber sido específicamente programada para ello con anterioridad.

La automatización no es algo nuevo, el concepto suele utilizarse en el ámbito de la industria con referencia al sistema que permite que una máquina desarrolle ciertos procesos o realice tareas sin intervención del ser humano. La automatización permite ahorrar tiempo y, muchas veces, dinero. La automatización puede observarse en diversas actividades que anteriormente eran realizadas por personas como: porteros, operarios en líneas de producción, responsables de caja en supermercados o peajes, operarios de almacenes logísticos, entre otros. Es muy común referirse a la automatización y a la inteligencia artificial de manera indistinta, sin embargo, no son lo mismo, en general ambas permiten realizar de forma automática tareas que tradicionalmente desempeñaba un humano, pero la principal diferencia entre inteligencia artificial y automatización es que:

- La automatización utiliza un software que sigue unas reglas y unos pasos preprogramados.
- La inteligencia artificial es capaz de realizar tareas de manera más “inteligente” y tomar decisiones o realizar acciones para las que no ha sido específicamente programada previamente.

3.4 Aprendizaje automático

Desde que nacen hasta que mueren los seres humanos llevan a cabo diferentes procesos, entre ellos se encuentra el del aprendizaje, por medio del cual adquieren conocimientos y desarrollan habilidades para poder analizar y entender el mundo que los rodea. El aprendizaje en las personas se lleva a cabo de manera automática ya que es un proceso tan sencillo que pasa desapercibido cómo se realiza y todo lo que implica.

A diferencia de los seres humanos, a las máquinas hay que indicarles que hacer y como aprender de los datos que se les proporcionan para que en base a la experiencia puedan obtener mejores resultados.

El aprendizaje automático o aprendizaje de máquinas (*machine learning* en inglés) es una rama de la inteligencia artificial que busca desarrollar técnicas para propiciar el aprendizaje de las computadoras. En otras palabras, sería encontrar la forma de transformar datos en información que entienda y pueda utilizar la computadora para realizar alguna acción para la cual no haya sido previamente programada explícitamente.

De acuerdo con Russell y Norvig “*una maquina aprende cuando mejora su desempeño con base en la experiencia*” (Russell, J., & Norvig, 2009). Es decir que entre más datos se tengan disponibles se obtendrían mejores resultados, los cuales deberían ser capaces de generalizar comportamientos para un conjunto más amplio de datos.

En su forma más básica, el aprendizaje automático utiliza algoritmos programados que reciben y analizan datos de entrada para predecir los valores de salida dentro de un rango aceptable. A medida que se introducen nuevos datos en estos algoritmos, aprenden y optimizan sus operaciones para mejorar el rendimiento, desarrollando “*inteligencia*” con el tiempo (APD, 2019).

3.4.1 Modelos

Una vez que se lleva a cabo el proceso de aprendizaje automático se pueden obtener modelos que pueden resolver el problema planteado, algunos modelos son:

- **Modelos geométricos**, son aquellas en donde las observaciones o datos pueden ser representados en un plano, los cuales pueden ser separables mediante una línea recta.
- **Modelos probabilísticos**, son aquellos modelos que buscan encontrar una relación entre un conjunto de atributos y la clase objetivo. Estos modelos se basan en el uso de la estadística bayesiana.
- **Modelos lógicos**, son aquellos modelos que representan reglas en forma de árboles.

3.4.2 Clases de Algoritmos

Los algoritmos de aprendizaje automático se agrupan de acuerdo al resultado de los mismos. La clasificación más general es:

- **Aprendizaje supervisado**, este tipo de algoritmo como su nombre lo indica es entrenado previamente con datos que ya han sido identificados y etiquetados. Una vez entrenado el algoritmo arroja resultados los cuales llegan a ser más precisos en base al número de datos con que se haya entrenado el algoritmo.
- **Aprendizaje no supervisado**, este tipo de algoritmo no se alimenta con datos de entrenamiento, sino que analiza los datos obtenidos y busca similitudes

entre los datos proporcionados para crear grupos de registros con características similares, de modo que puede clasificar los nuevos registros en el grupo en el que el grado de similitud sea mayor.

- **Aprendizaje por refuerzo**, este tipo de aprendizaje va mejorando conforme se utiliza ya que recibe una retroalimentación por cada resultado obtenido, el cual se almacena en su memoria para utilizar esa “experiencia” posteriormente.

3.4.3 Tipos de Algoritmos

Los problemas de aprendizaje automático pueden resolverse mediante el uso de uno o más algoritmos, dependiendo cual sea el problema que se presente, algunos de los algoritmos más representativos son:

- **Algoritmos bayesianos**, están fundamentados en el teorema de Bayes y trabajan con el supuesto de que cada valor es independiente de los demás con lo que puede predecir una clase en base al conjunto de datos utilizando la probabilidad. Es uno de clasificadores con mejores resultados ya que es fácil de implementar y sus resultados son más que aceptables llegando a competir con otros métodos de clasificación más sofisticados.
- **Los algoritmos de regresión**, tratan de comprender la relación existente entre las variables dependiente e independiente. Este tipo de algoritmos son mayormente utilizados para predecir sucesos en base a datos históricos.
- **Algoritmos de agrupación**, este algoritmo sirve para categorizar datos que no han sido etiquetados previamente, el cual lleva a cabo una búsqueda iterativa de grupos dentro del conjunto de registros con la que puede asignar cada punto a cada uno de los grupos encontrados.
- **Algoritmos de árbol de decisión**, este algoritmo representa cada uno de los resultados posibles mediante un diagrama similar a un árbol, en donde cada

nodo representa una condición aplicada a la variable o característica y cada rama representa las opciones disponibles.

- **Algoritmos de redes neuronales**, una red neuronal artificial (RNA) esta inspirada en los sistemas biológicos, como el cerebro, y en cómo este procesa la información mediante las neuronas, es por esto que trata de representar cada una de estas neuronas mediante funciones interconectadas, las cuales trabajan juntas para resolver el problema planteado. Este tipo de algoritmo es bastante útil para problemas no lineales o donde la relación entre las variables no es fácilmente comprensible.

3.5 Minería de datos

De acuerdo a la definición de Maimon y Rokach “*La minería de datos, es un campo de la estadística y las ciencias de la computación referido al proceso que intenta descubrir patrones en grandes volúmenes de conjuntos de datos*” (Maimon & Rokach, 2010). Utiliza los métodos de la inteligencia artificial, aprendizaje automático, estadística y sistemas de bases de datos.

Según la definición de Microsoft, la minería de datos utiliza los principios estadísticos para detectar patrones o clasificar los datos. El uso de algoritmos de minería de datos permite realizar predicciones, encontrar patrones, realizar recomendaciones y descubrir nueva información que no puede apreciarse a simple vista.

La minería de datos es el proceso de detectar la información procesable de los conjuntos grandes de datos. Utiliza el análisis matemático para deducir los patrones y tendencias que existen en los datos (Microsoft, 2019). Algunos de los escenarios en donde pueden aplicarse los modelos de minería de datos son los siguiente:

- **Previsión:** Estimación de ventas, predicción de cargas de servidor o tiempo de inactividad del servidor.
- **Riesgo y probabilidad:** Elección de los mejores clientes para el envío de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, asignación de probabilidades a diagnósticos u otros resultados.
- **Recomendaciones:** Determinación de los productos que es probable que se vendan juntos, generación de recomendaciones.
- **Buscar secuencias:** Análisis de las selecciones de los clientes en un carro de la compra y predicción de los siguientes eventos probables
- **Agrupación:** Separación de los clientes o eventos en el clúster de elementos relacionados, análisis y predicción de afinidades.

3.6 Descubrimiento de conocimiento en bases de datos (KDD)

Al manejar grandes cantidades de datos, el descubrimiento de conocimiento en bases de datos o KDD se refiere al proceso de identificar patrones válidos, novedosos, potencialmente útiles y principalmente entendibles (Figura 5).

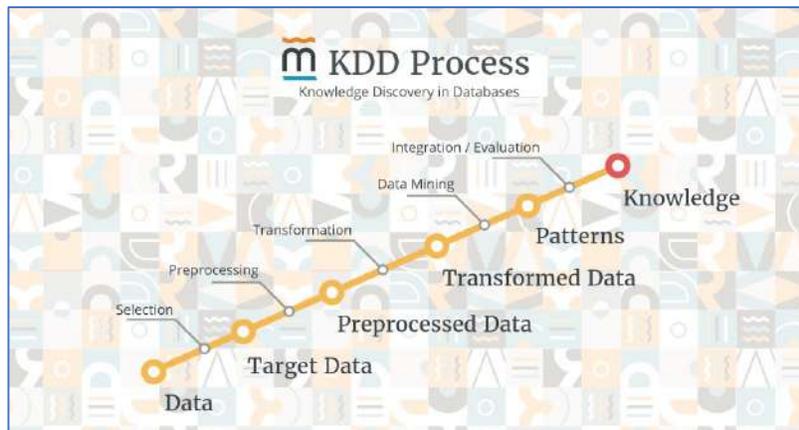


Figura 5 Proceso de Obtención de Conocimiento (KDD)

Knowledge Discovery implica la evaluación e interpretación de patrones y modelos para tomar decisiones con respecto a lo que constituye conocimiento y lo que no lo es. Por lo tanto, el KDD requiere de un amplio y profundo conocimiento sobre el área de estudio con la que se esté trabajando. Por otra parte, la Minería de Datos, exploración de datos o Data Mining, no requiere tanto conocimiento sobre el área de estudio, sino más conocimiento técnico.

Como se mencionó anteriormente, la Minería de Datos es un paso que forma parte del KDD e implica el análisis de grandes cantidades de datos observacionales, para encontrar relaciones no observables, es decir, el Data Mining se ocupa de reunir los datos de manera novedosa, entendible y útil para el propietario o usuario final.

3.6.1 Etapas del KDD

El proceso se compone de 6 etapas fundamentales, las cuales se describen a continuación:

- 1. Comprensión del dominio del estudio y establecimiento de objetivos**, en esta etapa se definen los límites y objetivos del proceso, así como también se determinan las fuentes de información más importantes que utilizarán.
- 2. Creación de un dataset objetivo**, se seleccionan y se integran los datos provenientes de distintas fuentes. En esta etapa es importante tratar de homogenizar los datos con el fin de facilitar su análisis y procesamiento.
- 3. Limpieza y procesamiento de datos**, busca mejorar la calidad de los datos con la finalidad de obtener resultados más aceptables.
- 4. Minería de datos**, en esta etapa es donde se aplica uno o más algoritmos a los datos con el fin de obtener información de utilidad.
- 5. Interpretación de los patrones minados**, se analizan los resultados obtenidos de los algoritmos, en esta etapa se obtienen los modelos de los datos

que fueron analizados. Los resultados obtenidos deben presentarse de una forma que sea entendible para quien utilizará la información.

- 6. Utilización del conocimiento descubierto**, esta es la última etapa del proceso, aquí se utiliza la información obtenida para tomar decisiones o plantear alguna estrategia para la mejora de algún proceso en la empresa donde se haya implementado.

3.7 Naive Bayes y el problema de la clasificación

La clasificación es un tipo de problema en el aprendizaje automático donde se busca encontrar patrones para predecir una agrupación de elementos o registros. La clasificación puede ser supervisada o no supervisada dependiendo el problema. En el caso de la supervisada, los patrones se obtienen basándose en un conjunto de datos previamente identificados y etiquetados, mientras que en el caso de la no supervisada se denomina como *clustering* en el cual los datos se agrupan en base a alguna medida de similitud entre los elementos. Un ejemplo son los sistemas de correo electrónico, los cuales pueden determinar si un correo es spam o no lo es, en base al análisis de los datos del correo, como palabras clave, IP, etc.

Algunos de los elementos que están presentes durante la clasificación son:

- Clasificador, son los algoritmos que implementan la clasificación.
- Variables, son los atributos de los elementos a clasificar.
- Clases, la categoría que se busca predecir.

En teoría de la probabilidad y minería de datos, un clasificador Bayesiano es un clasificador probabilístico fundamentado en el teorema de Bayes (Han, 2011), se utiliza en el modelado de predicción y de exploración. Si una nueva tupla va a ser clasificada, el Teorema de Bayes es utilizado para calcular la probabilidad de que

pertenezca a la clase utilizando la ecuación (1) (Mozina, Demsar, Kattan, & Zupan, 2004). En la ecuación (1) la P denota probabilidad y la notación $P(T|C_i)$ representa la probabilidad condicional de T dado que se sabe que la clase es C_i . C_i es una clase que pertenece al conjunto de clases $\{C_1, C_2, C_3, \dots\}$ para el conjunto de datos. La ecuación (1) es calculada para cada una de las clases y la clase con $P(C_i|T)$ más alta es considerada como instancia de la clase.

$$P(C_i|T) = \frac{P(T|C_i)P(C_i)}{P(T)} \quad (1)$$

Al calcular $P(C_i|T)$ para cada C_i el denominador $P(T)$ es constante en todas las clases por lo tanto puede ser eliminado de la ecuación. Así, la ecuación (2) se puede usar para encontrar la clase que tiene la mayor probabilidad. El símbolo \sim denota que el lado izquierdo es proporcional al lado derecho.

$$P(C_i|T) \sim P(T|C_i)P(C_i) \quad (2)$$

El clasificador Naive Bayes se basa en el supuesto de independencia condicional de clase, es decir, que los valores de los atributos de T son independientes entre sí. Como consecuencia, si T es la tupla $n \langle t_1, t_2, \dots, t_n \rangle$, entonces $P(T|C_i)$ de la ecuación (2) puede ser calculada utilizando la ecuación (3). La justificación de la ecuación (3) se basa en la teoría de probabilidad, donde la probabilidad conjunta de eventos independientes se puede calcular multiplicando las probabilidades de estos eventos. Por lo tanto, para calcular $P(C_i|T)$ basado en la ecuación (2) necesitamos calcular $P(C_i)$ y calcular $P(T|C_i)$ basados en Ecuación (3) y multiplica los dos resultados. Esto se realiza para cada clase C_i y la clase con el valor más alto es elegido como la clase para la nueva tupla T.

$$P(T|C_i) = \prod_{k=1}^n P(t_k|C_i) = P(t_1|C_i) \times P(t_2|C_i) \times P(t_3|C_i) \times \dots \times P(t_n|C_i) \quad (3)$$

3.9 Python y el análisis de datos

Python es un lenguaje de programación interpretado cuya filosofía hace hincapié en la legibilidad de su código. 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, dinámico y multiplataforma.

Python es multipropósito, se lo puede utilizar en ciencias de datos, desarrollo y administración de sistemas, construcción de aplicaciones web y scripts, entre otros. Contiene paquetes (“packages”) que ayudan en las tareas de ciencias de datos, por ejemplo, para hacer el análisis exploratorio (EDA exploratory data analysis) y trabajar con inteligencia artificial (redes neuronales y deep learning). Pero por ahora, solo Python puede utilizar TensorFlow.

Python utiliza un software adicional para el manejo de paquetes como Anaconda, Miniconda, or PIP. Es administrado por la Python Software Foundation. Posee una licencia de código abierto, denominada Python Software Foundation License,² que es compatible con la Licencia pública general de GNU a partir de la versión 2.1.1, e incompatible en ciertas versiones anteriores.

3.9.1 Django framework

Django es un framework de desarrollo web de código abierto, escrito en Python, que respeta el patrón de diseño conocido como Modelo–vista–template. Fue desarrollado en origen para gestionar varias páginas orientadas a noticias de la World Company de Lawrence, Kansas, y fue liberada al público bajo una licencia BSD en julio de 2005; el framework fue nombrado en alusión al guitarrista de jazz gitano

Django Reinhardt. En junio de 2008 fue anunciado que la recién formada Django Software Foundation se haría cargo de Django en el futuro.

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.

3.9.2 Scikit Learn y Pandas

Scikit-learn (también conocido como SKLearn) es una biblioteca de aprendizaje automático de software libre para el lenguaje de programación Python. Cuenta con varios algoritmos de clasificación, regresión y agrupamiento que incluyen máquinas de vectores de soporte, bosques aleatorios, aumento de gradiente, k- medias y DBSCAN, y está diseñado para interactuar con las bibliotecas numéricas y científicas de Python NumPy y SciPy.

Scikit-learn está escrito en gran parte en Python, y utiliza NumPy ampliamente para álgebra lineal de alto rendimiento y operaciones de matriz. Además, algunos algoritmos centrales están escritos en Cython para mejorar el rendimiento. Las máquinas de vectores de soporte se implementan mediante un envoltorio de Cython alrededor de LIBSVM; Regresión logística y máquinas de vectores de soporte lineal por un contenedor similar alrededor de LIBLINEAR.

Scikit-learn se integra bien con muchas otras bibliotecas de Python, como Matplotlib y Plotly para trazar, NumPy para vectorización de matriz, marcos de datos de pandas, SciPy, entre otras.



Figura 6 Librería pandas

Pandas es una librería de Python especializada en el manejo y análisis de estructuras de datos (Figura 6). Las principales características de esta librería son:

- Define nuevas estructuras de datos basadas en los arreglos de la librería NumPy pero con nuevas funcionalidades.
- Permite leer y escribir fácilmente ficheros en formato CSV, Excel y bases de datos SQL.
- Permite acceder a los datos mediante índices o nombres para filas y columnas.
- Ofrece métodos para reordenar, dividir y combinar conjuntos de datos.
- Permite trabajar con series temporales.
- Realiza todas estas operaciones de manera muy eficiente.

3.9.3 Naive Bayes con Scikit Learn

La librería Scikit Learn permite aplicar diferentes algoritmos de aprendizaje automático, entre ellos Naive Bayes el cual es el algoritmo elegido para este proyecto. Para utilizarla dentro del proyecto, deben importarse los modulos necesarios al inicio de nuestro archivo, los más característicos con el modulo GaussianNB el cual permite aplicar Naive Bayes mediante una distribución Gaussiana, por otro lado también se tienen `train_test_split` el cual se utiliza para separar los datos de prueba para el entrenamiento del modelo (Figura 7).

```
>>> from sklearn.datasets import load_iris
>>> from sklearn.model_selection import train_test_split
>>> from sklearn.naive_bayes import GaussianNB
>>> X, y = load_iris(return_X_y=True)
>>> X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.5, random_state=0)
>>> gnb = GaussianNB()
>>> y_pred = gnb.fit(X_train, y_train).predict(X_test)
>>> print("Number of mislabeled points out of a total %d points : %d"
...      % (X_test.shape[0], (y_test != y_pred).sum()))
Number of mislabeled points out of a total 75 points : 4
```

Figura 7 Aplicación del algoritmo Naive Bayes Gaussiano con la librería de Scikit Learn

Algunos de los métodos disponibles en la librería pueden observarse en el cuadro siguiente.

<code>fit(x, y[, sample_weight])</code>	Ajusta Naive Bayes Gaussiano de acuerdo con X, y
<code>get_params([deep])</code>	Obtiene los parámetros para este estimador.
<code>partial_fit(X, y[, classes, sample_weight])</code>	Ajuste incremental en un lote de muestras.
<code>predict(X)</code>	Realiza la clasificación en una matriz de vectores de prueba X.
<code>predict_log_proba(X)</code>	Devuelve estimaciones logarítmicas de probabilidad para el vector de prueba X.
<code>predict_proba(X)</code>	Devuelve estimaciones de probabilidad para el vector de prueba X.
<code>score(X, y[, sample_weight])</code>	Devuelve la precisión media en las etiquetas y los datos de prueba dados.
<code>set_params(**params)</code>	Define los parámetros de este estimador.

Tabla 1 Métodos disponibles para la librería GaussianNB

Capítulo 4. Metodología

La metodología utilizada para este proyecto está basada en el proceso de descubrimiento de conocimiento en bases de datos (KDD) el cual, como se menciona en el apartado 3.7 del capítulo del marco teórico, es utilizado en proyectos con grandes cantidades de datos para identificar patrones válidos, novedosos, potencialmente útiles y principalmente entendibles.

Tomando como referencia este proceso para la obtención de conocimiento se ha planteado la siguiente metodología (Figura 8), la cual ha sido adaptada para ajustarse a los objetivos del proyecto. Esta metodología constará de 5 etapas, las cuales son descritas a continuación.

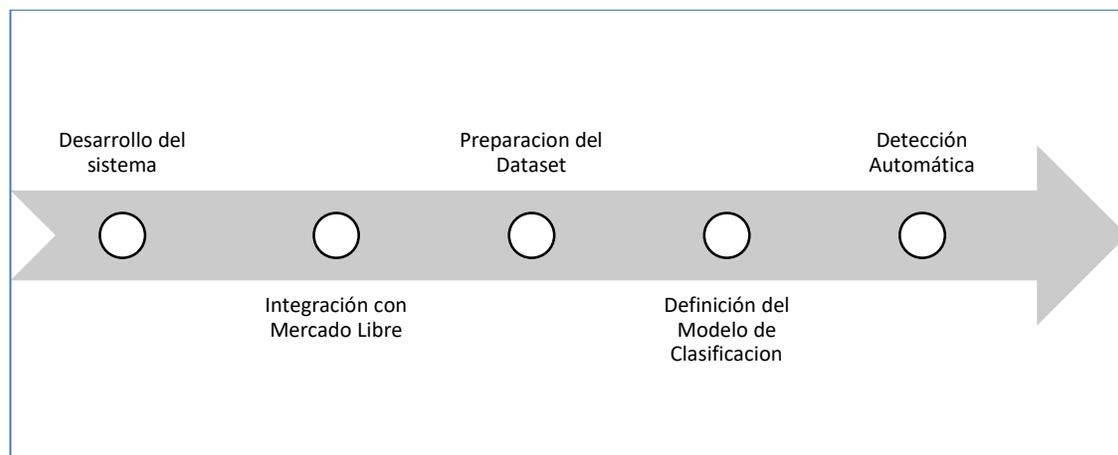


Figura 8 Metodología de Desarrollo

4.1 Desarrollo del sistema

Para la primera etapa del proyecto se pretende desarrollar un sistema que pueda integrarse con la plataforma de comercio electrónico Mercado Libre (Figura 9), con la finalidad de consultar y obtener los datos de vendedores y productos, para así poder analizarlos y determinar mediante un algoritmo de inteligencia artificial, el cual será abordado en los capítulos siguientes, mediante este algoritmo se determina si los productos analizados pueden considerarse como competencia del usuario y así monitorearlos para tomar decisiones.



Figura 9 Plataforma de Comercio Electrónico Mercado Libre

El desarrollo del sistema está pensado en seguir un enfoque web debido a los requerimientos del usuario y al requerir la integración de librerías de aprendizaje automático se ha decidido utilizar el lenguaje Python con ayuda del framework Django, el cual permite un desarrollo rápido y una gestión eficiente de los cambios al sistema. Además de que es un software de licencia libre por lo que no generaría costos adicionales al utilizarlo (Figura 10).



Figura 10 Django Framework

4.1.1 Diagrama de casos de uso

La aplicación está desarrollada en esta primera versión para ser utilizada por un solo usuario, el cual tendrá acceso a las siguientes funcionalidades:

- Acceder a la aplicación
- Vincular su cuenta de MercadoLibre
- Importar su catalogo
- Búsqueda de Productos
- Detección y seguimiento de competencia
- Reporte de seguimiento de productos
- Sugerencias y actualización de precios

En la figura 11 se representa mediante un diagrama de casos de uso las funcionalidades anteriores.

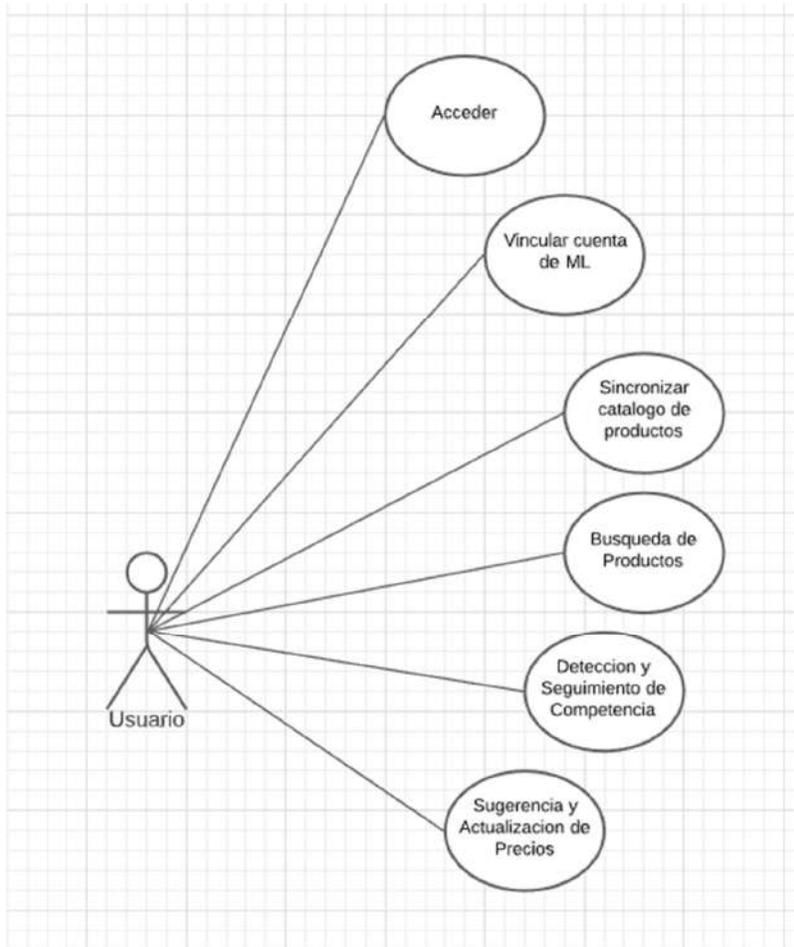


Figura 11 Diagrama de Casos de Uso

4.1.2 Diagrama de clase

Para el proyecto MLCOMPET se ha realizado un análisis de los datos que se van a utilizar y almacenar, para ellos se ha planteado el siguiente diagrama de clases en donde se utilizan los modelos siguientes:

- **Producto:** Es el modelo principal que se va a utilizar, en este modelo se almacena la información de los productos del usuario y de otros vendedores, en él se guarda el nombre, el precio, el vendedor, sku, entre otros datos.
- **Vendedor:** Es el modelo en donde se guarda la información de los vendedores, se guarda el nombre, su puntuación, nickname, entre otros datos.

- **Búsqueda:** Este modelo se pensó para guardar las búsquedas realizadas por el usuario y ejecutarlas mas rápido guardando la información en la memoria temporal del navegador.
- **Seguimiento:** En este modelo se almacenan los productos que se desean monitorear para poder actualizar los datos de los productos en caso de que el competidor ofrezca un mejor precio.

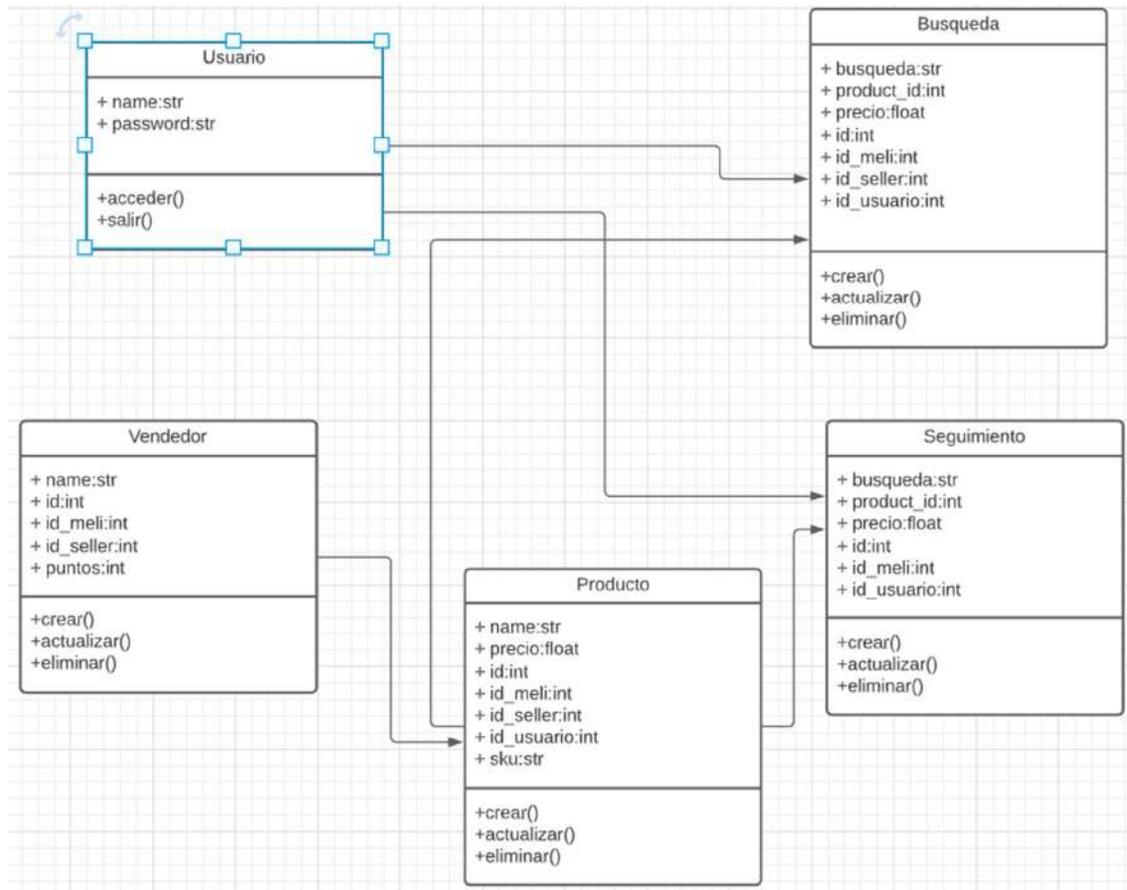


Figura 12 Diagrama de clases

4.2 Integración con Mercado Libre

Para integrar el sistema con Mercado Libre se ha planteado hacerlo a través del API REST de la plataforma, la cual como se mencionó en capítulos anteriores se trata de una interfaz para la programación de aplicaciones, haciendo uso de los distintos métodos HTTP, los cuales se utilizan comúnmente en el desarrollo de aplicaciones para consumir o enviar recursos hacia otras plataformas o desarrollos. Mercado Libre expone un API REST pública que permite consultar e interactuar con los datos generados y almacenados en la plataforma, para así poder crear nuevas aplicaciones. De acuerdo a la documentación, para integrar el sistema con la plataforma se deben seguir las etapas del proceso de integración (Figura 13).

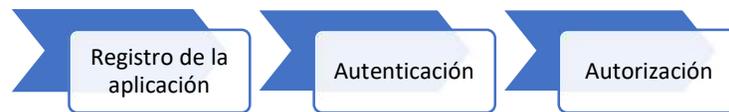


Figura 13 Proceso de Integración con Mercado Libre

4.2.1 Registro de la Aplicación

El primer paso de la integración consiste en registrar una aplicación dentro de la plataforma de Mercado Libre, para ello se debe contar con una cuenta activa usuario y acceder al área de desarrollo llamada *Mercado Libre Developers* en donde se encuentran todos los recursos necesarios para realizar la integración. Para acceder, se debe entrar a la siguiente URL <https://developers.mercadolibre.com/> (Figura 14) y seleccionar el país correspondiente al lugar donde estamos registrando el desarrollo, en este caso México.



Figura 14 Plataforma de Desarrollo de Mercado Libre

El siguiente paso es registrar una nueva aplicación en el administrador de aplicaciones de la plataforma (Figura 15). Para ello, se deben proporcionar algunos datos sobre el desarrollo a realizar.

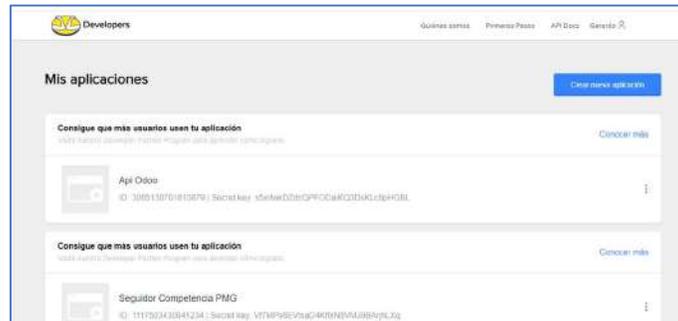


Figura 15 Administrador de Aplicaciones de Mercado Libre

Algunos de los datos que se solicitan son el nombre de la aplicación, una breve descripción de la misma, así como también una imagen de la aplicación, dichos datos deben capturarse en el formulario que se presenta (Figura 16).

The image shows a form titled 'Información básica' for creating a new application. It contains several input fields: 'Nombre*' and 'Nombre corto*' (both with a circular refresh icon), 'Descripción*' (with a circular refresh icon), and 'Logo'. Below the logo field, there is a note: 'Formato PNG de 400x400 pixels.' and a blue plus icon in a square box, indicating where to upload the logo image.

Figura 16 Formulario de Información Básica de la Aplicación

El siguiente paso es indicar la URL de la aplicación desarrollada, así como el alcance y las áreas que va a tener la aplicación, para ello se seleccionan los recursos a los que va a tener acceso y el nivel de permisos para cada uno ellos.

Figura 17 Formulario de Alcance y Permisos de la Aplicación

Una vez registrada la aplicación esta debe aparecer en el administrador de aplicaciones de la plataforma (Figura 18), la cual debe mostrar los siguientes datos, los cuales se van a utilizar para conectar la aplicación a la plataforma.

- **App_id:** Identificador único de la aplicación.
- **Secret_key:** Clave única de la aplicación utilizado en conjunto con la *app_id* para obtener un Access token para conectarse a la aplicación.

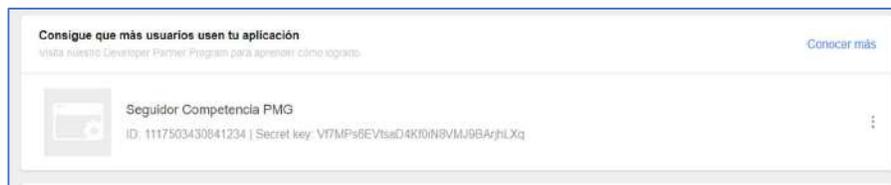


Figura 18 Administrador de Aplicaciones de Mercado Libre

4.2.2 Autenticación

El proceso de autenticación y autorización permiten comenzar a utilizar los recursos de la plataforma, para ellos es necesario desarrollar los procesos de autenticación y autorización dentro de la aplicación desarrollada, de esta manera, se podrá trabajar con los recursos privados del usuario cuando autorice el acceso a la aplicación.

El proceso de autenticación es utilizado para verificar la identidad de una persona en función de uno o varios factores, asegurando que los datos de quien los envió sean los correctos. Los pasos a seguir son:



1. Autenticarse con el usuario de Mercado Libre.

El primer paso se solicita que el usuario ingrese a su cuenta de Mercado Libre a la cual se va a conectar (Figura 19).

Figura 19 Autenticación del usuario en Mercado Libre

2. Obtener Autorización

La URL <http://auth.mercadolibre.com.mx/authorization> permite obtener la autorización del usuario para que se puedan consultar sus datos por medio de la aplicación. Para ello se debe proporcionar el id de la aplicación registrada previamente, así como la URL a la cual va a redirigir una vez hecha la autorización.

En este punto es importante indicar que la URL debe ser la misma que se registró previamente en el administrador de aplicaciones ya que en caso contrario la autenticación no será realizada.

3. Vincular la aplicación con la cuenta del usuario

En el último paso, el usuario es redirigido a la pantalla donde se le solicitará que vincule la aplicación con su cuenta (Figura 20).

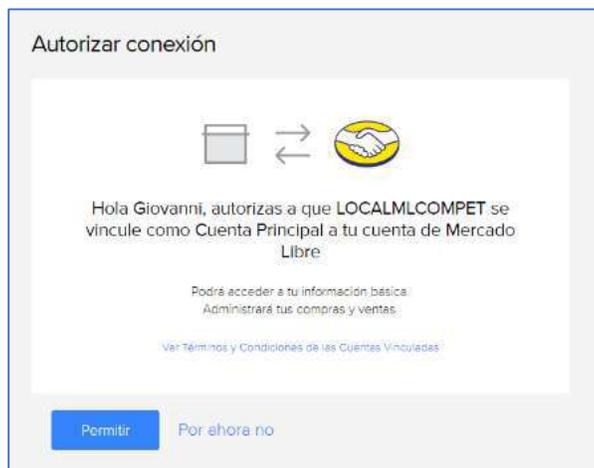


Figura 20 Autorización de acceso a Mercado Libre

Al revisar la URL se puede observar que se agregó el parámetro CODE, el cual será utilizado cuando se necesite generar un ACCESS_TOKEN que permitirá el acceso al API (Figura 21).



Figura 21 Parámetro Code agregado a la URL de Autorización

4.2.3 Autorización

La autorización es el proceso por el cual se le da permiso a los usuarios para acceder a recursos privados. En este proceso, se debe definir qué recursos y operaciones se pueden realizar (“sólo lectura” o “lectura y escritura”). Esto se logra a través del protocolo OAuth 2.0, uno de los uno de los más utilizados en plataformas abiertas como Twitter y Facebook ya que es un método seguro para trabajar con recursos privados. Este protocolo brinda las siguientes características:

- **Confidencialidad**, el usuario no deberá revelar su clave en ningún momento.
- **Integridad**, solo podrán ver datos privados aquellas aplicaciones que tengan el permiso de hacerlo.
- **Disponibilidad**, los datos siempre estarán disponibles en el momento que se necesiten.

Una vez que el usuario se ha autenticado y ha autorizado la aplicación para acceder a los datos de su cuenta, se obtiene un código de autorización, el cual se utiliza posteriormente para obtener un *ACCESS_TOKEN* el cual permite que la aplicación pueda interactuar con Mercado Libre y tener acceso a los recursos privados del usuario como ventas, publicaciones, productos, datos del vendedor, calificaciones, entre otros.

Para solicitar un **ACCESS_TOKEN** se debe realizar una petición *HTTP POST* a la URL <https://api.mercadolibre.com/oauth/token> proporcionando los siguientes parámetros:

- **Auth Code**: indica que la operación actual es intercambiar un código de autorización por un token de acceso.
- **App Id**: Es el Id de la aplicación registrada previamente en Mercado Libre.
- **Secret Key**: Es la clave secreta obtenida previamente al registrar la aplicación en Mercado Libre
- **Code**: Es el Código de autorización obtenido al momento que el usuario autoriza el acceso a la aplicación.
- **Redirect Url**: Url a la cual el usuario va a ser redirigido al obtener el *ACCESS_TOKEN*.

```
{
  "access_token": "APP_USR-5387223166827464-090515-8cc4448aac10d51054",
  "token_type": "bearer",
  "expires_in": 10800,
  "scope": "offline_access read write",
  "user_id": 8035443,
  "refresh_token": "TG-5b9032b4e4b0714aed1f959f-8035443"
}
```

Figura 22 Ejemplo de Access Token obtenido de Mercado Libre

En la respuesta obtenida se puede observar el tiempo de expiración, el id del usuario que lo genero, los permisos, entre otros datos (Figura 22). El *ACCESS_TOKEN* generado tiene una duración de 6 horas desde que se solicitó, una vez transcurrido ese periodo es necesario volver a solicitar autorización, en caso de requerir seguir trabajando de manera prolongada con la aplicación se utiliza el *REFRESH_TOKEN* el cual permite actualizar el *ACCESS_TOKEN* sin volver a solicitar autorización por parte del usuario.

4.2.4 Flujo de Conexión con Mercado Libre

De acuerdo a todo lo mencionado anteriormente, a continuación, se representa de manera general como es flujo de comunicación entre la aplicación y Mercado Libre, con el cual se obtiene un *ACCESS_TOKEN* mediante el cual se consumen los recursos de la plataforma (Figura 23).

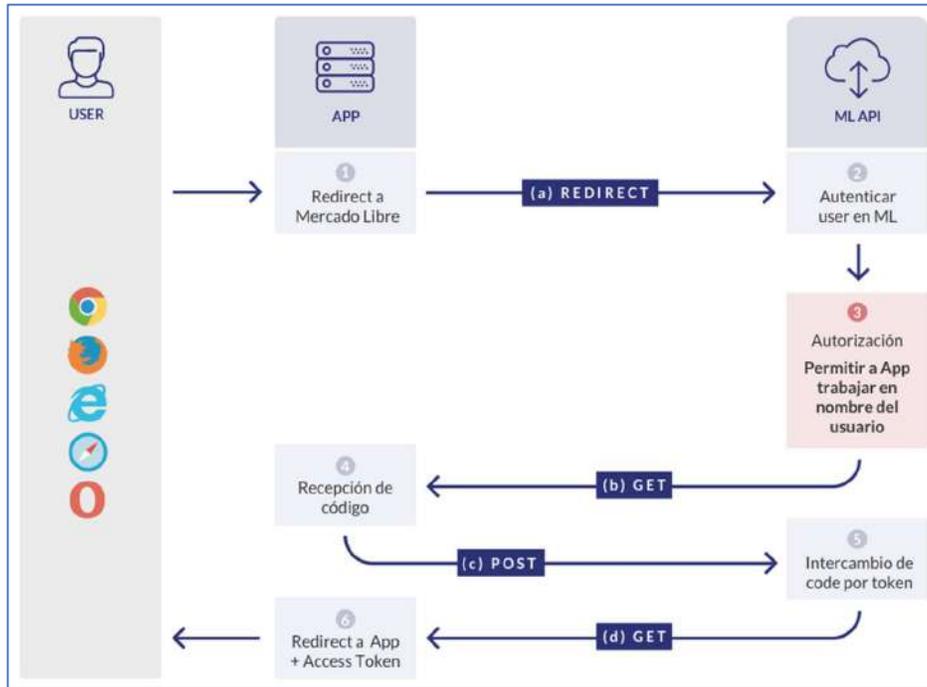


Figura 23 Flujo de Conexión a Mercado Libre

1. Redirecciona la aplicación a Mercado Libre.
2. Autenticación en Mercado Libre
3. Página de autorización.
4. POST para intercambiar el código de autorización por un *ACCESS_TOKEN*.
5. La API de Mercado Libre intercambia el código de autorización por un *ACCESS_TOKEN*
6. El *ACCESS_TOKEN* obtenido ya se puede utilizar para realizar llamadas al API y así obtener acceso a los datos privados del usuario.

4.2.5 Obtención de datos del usuario

Como se ha mencionó en los apartados anteriores, los recursos de Mercado Libre se encuentran disponibles a través de su API REST y se puede acceder a ellos utilizando diferentes tecnologías, ya sea tecnologías del lado del cliente como JavaScript o del lado del servidor como Python o Java respectivamente.

El API REST de Mercado Libre ofrece distintos recursos que pueden ser utilizados y consumidos por los desarrolladores. Entre los recursos disponibles se encuentran:

- **Usuarios y Aplicaciones:** permite trabajar con usuarios y aplicaciones registradas en Mercado Libre.
- **Categorías y Publicaciones:** permite utilizar las opciones del árbol de categorías y listas de Mercado Libre.
- **Ubicación y Monedas:** Se utiliza para obtener las ubicaciones y monedas disponibles.
- **Productos y Búsquedas:** Muestra los datos relacionados a los productos publicados.
- **Preguntas y Respuesta:** Permite gestionar las preguntas y respuestas realizadas por los usuarios.
- **Pedidos y Opiniones:** Permite gestionar los pedidos y las opiniones de los usuarios.
- **Métricas:** Permite conocer algunas métricas de la plataforma como número de visitas de un usuario, total de preguntas, numero de clic, entre otras.
- **Envíos:** Permite gestionar los recursos necesarios para realizar los envíos como datos de envío, métodos disponibles, costos, entre otros.

Para esta etapa del proyecto se pretende obtener los datos de productos y usuarios desde la plataforma, para ello se deben utilizar los recursos de **Items y Búsquedas**, así como también, el recurso de **Usuarios y Aplicaciones** los cuales cuentan con un conjunto amplio de métodos que permitirán trabajar con los datos disponibles en la plataforma. A continuación, se explican los métodos disponibles para los recursos mencionados.

4.2.6 Recurso de productos y búsquedas

El recurso de productos y búsqueda cuenta con diferentes métodos los cuales se revisarán a continuación, los cuales permitirán obtener los datos necesarios para utilizarlos en este proyecto (Figura 24).

Recurso	Descripción
/sites/\$SITE_ID/search?category=\$CATEGORY_ID	Obtener ítems listados en una categoría.
/sites/\$SITE_ID/search?q=Motorola%20G6	Obtener ítems de una consulta de búsqueda.
/sites/\$SITE_ID/search?nickname=\$NICKNAME	Obtener ítems de los listados por nickname.
/sites/\$SITE_ID/search?seller_id=\$SELLER_ID	Permite listar ítems por vendedor.
/sites/\$SITE_ID/search?seller_id=\$SELLER_ID&category=\$CATEGORY_ID	Obtener ítems de los listados por vendedor en una categoría específica.
/users/\$USER_ID/items/search?access_token=\$ACCESS_TOKEN	Permite listar todos los ítems de la cuenta de un vendedor.

Figura 24 Métodos disponibles para el recurso de Items y Búsquedas

Categorías de Productos

En Mercado Libre, las categorías son un conjunto jerárquico de grupos en los cuales se enumeran las publicaciones de naturaleza similar, denominados *Árbol de Categorías*. Estas categorías ayudan a los usuarios a buscar fácilmente el tipo de publicación que desean y cada sitio (Argentina, Brasil, México, etc.) tiene su propio conjunto de categorías (Mercado Libre Developers, 2020). Las categorías se registran por cada país, por lo tanto, para obtener las categorías disponibles en la plataforma para un país en particular se debe proporcionar el código del país del cual queremos consultar, en este caso se utiliza el código MLM que corresponde a México. Para consultar las categorías disponibles para México se haría la siguiente llamada:

```
curl -X GET https://api.mercadolibre.com/sites/MLM/categories
```

Para realizar las llamadas al API REST se está utilizando el comando curl de la consola de comandos del sistema (CMD), el cual permite realizar llamadas HTTP a otros servidores. La llamada anterior devuelve una respuesta, la cual está en un formato JSON el cual como se comentó en apartados anteriores se utiliza para compartir información entre aplicaciones y se ha adoptado como estándar al utilizar API de tipo REST.

4.2.7 Búsqueda de productos por Categoría

Para consultar los productos que pertenecen a determinada categoría se debe realizar la llamada siguiente, proporcionando los parámetros `$site_id` y `$category_id` en donde el primero corresponde al código del país y segundo corresponde a la categoría.

```
curl -X GET https://api.mercadolibre.com/sites/\$SITE\_ID/search?category=\$CATEGORY\_ID
```

Por ejemplo, para consultar los productos que pertenecen a la categoría de telefonía y celulares se utilizaría la categoría MLM1051 de acuerdo a la consulta

realizada previamente al catálogo de categorías de productos para México. Entonces la llamada quedaría de la siguiente forma:

```
curl -X GET https://api.mercadolibre.com/sites/MLM/search?category=MLM1051
```

La respuesta obtenida por la llamada anterior muestra los resultados de las publicaciones encontradas para esa categoría, mostrando datos como el número de registros total, así como datos de cada publicación encontrada (Figura 25).

```
{
  "site_id": "MLA",
  "paging": {
    "total": 90591,
    "offset": 0,
    "limit": 50,
    "primary_results": 1039
  },
  "results": [
    {
      "id": "MLA811601010",
      "site_id": "MLA",
      "title": "Samsung Galaxy J4+ Dual Sim 32 Gb Negro (2 Gb Ram)",
      "seller": {
        "id": 451257294,
        "power_seller_status": null,
        "car_dealer": false,
        "real_estate_agency": false,
        "tags": []
      },
      "price": 19609,
      "currency_id": "ARS",
      "available_quantity": 1,
    }
  ]
}
```

Figura 25 Consulta de Productos por Categoría

4.2.8 Búsqueda de productos por palabra clave

Para consultar productos por nombre se debe proporcionar una o varias palabras clave las cuales permiten filtrar los productos que cumplan con ese criterio de búsqueda. Para hacer la llamada hacia el API REST se deben proporcionar los parámetros \$SITE_ID que corresponde al código del país, en este caso para México tomara el valor de MLM, y también proporcionar el parámetro \$CONSULTA el cual debe contener la consulta de búsqueda.

```
curl -X GET https://api.mercadolibre.com/sites/\$SITE\_ID/search?q=\$CONSULTA
```

Por ejemplo, para buscar los equipos celulares que sean Motorola G6 se realizaría la siguiente llamada.

```
curl -X GET https://api.mercadolibre.com/sites/MLM/search?q=Motorola%20G6
```

Para la consulta anterior obtendríamos la siguiente respuesta, en donde podemos observar algunos datos relevantes como el número total de registros obtenidos, los datos de las diferentes publicaciones que coincidieron con el criterio de búsqueda proporcionado, por ejemplo, el precio del producto publicado, los datos del vendedor, la categoría a la que pertenece, entre algunos otros datos que se utilizan más adelante (Figura 26).

```
{
  "site_id": "MLA",
  "query": "Motorola G6",
  "paging": {
    "total": 916,
    "offset": 0,
    "limit": 50,
    "primary_results": 916
  },
  "results": [
    {
      "id": "MLA810645375",
      "site_id": "MLA",
      "title": "Motorola G6 Plus 64 Gb Nimbus",
      "seller": {
        "id": 143125485,
        "power_seller_status": "platinum",
        "car_dealer": false,
        "real_estate_agency": false,
        "tags": []
      },
      "price": 17999,
    }
  ]
}
```

Figura 26 Respuesta de llamada a API REST de productos por palabra clave

4.3 Preparación del dataset

El siguiente paso en la metodología consiste en consolidar y limpiar los datos obtenidos de la etapa previa. Los datos pueden encontrarse distribuidos y almacenados en formatos distintos, también pueden contener incoherencias como datos faltantes o incorrectos (Figura 27).

La limpieza de datos no solamente implica quitar los datos no válidos o interpolar valores que faltan, sino también buscar las correlaciones ocultas en los datos, identificar los orígenes de datos que son más precisos y determinar qué columnas son las más adecuadas para el análisis. Por ejemplo, ¿debería utilizar la fecha de envío o la fecha de pedido? ¿Qué influye más en las ventas: ¿la cantidad, el precio total o un precio con descuento? Los datos incompletos, los datos incorrectos y las entradas que parecen independientes, pero que de hecho están estrechamente correlacionadas, pueden influir en los resultados del modelo de maneras que no espera. (Microsoft, 2019).

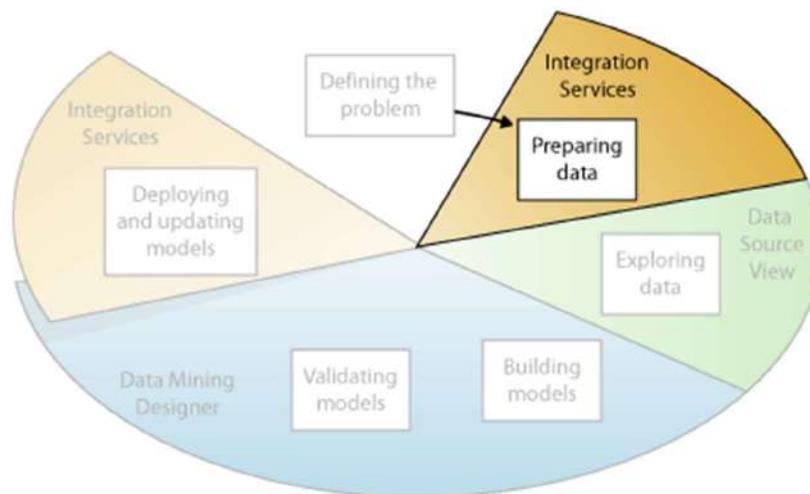


Figura 27 Preparación del Dataset

Como buena práctica se empieza con un dataset inicial para familiarizarse con los datos, encontrar así las primeras idpreeas sobre ellos y tener un buen entendimiento de cualquier posible problema de calidad de datos. La preparación de datos suele ser a menudo un proceso que consume mucho tiempo y muy propenso al error. El viejo dicho de “entra basura, sale basura” es muy aplicable a aquellos proyectos de ciencia de datos que reúnen datos con valores inválidos, fuera de rango, o simplemente inexistentes. Analizar datos que no fueron revisados cuidadosamente ante esos problemas puede producir resultados engañosos. Es así entonces que el éxito de un proyecto de ciencia de datos depende fuertemente de la calidad de los datos preparados (DataScience, 2018).

4.3.1 Exploración de datos

En esta etapa se deben conocer los datos para tomar las decisiones adecuadas al crear los modelos de minería de datos que serán utilizados en el proyecto. Entre las técnicas de exploración se incluyen calcular los valores mínimos y máximos, calcular la media y las desviaciones estándar, y examinar la distribución de los datos. Por ejemplo, al revisar el máximo, el mínimo y los valores de la media se podría determinar que los datos no son representativos de los clientes o procesos de negocio, y que por consiguiente debe obtener más datos equilibrados o revisar las suposiciones que son la base de sus expectativas. Las desviaciones estándar y otros valores de distribución pueden proporcionar información útil sobre la estabilidad y exactitud de los resultados. Una desviación estándar grande puede indicar que agregar más datos podría ayudarle a mejorar el modelo. Los datos que se desvían mucho de una distribución estándar se podrían sesgar o podrían representar una imagen precisa de un problema de la vida real, pero dificultar el ajustar un modelo a los datos.

Al explorar los datos para conocer el problema empresarial, puede decidir si el conjunto de datos contiene datos defectuosos y, a continuación, puede inventar una estrategia para corregir los problemas u obtener una descripción más profunda de los comportamientos que son típicos de su negocio (Microsoft, 2019).

4.3.2 Selección de datos

En esta etapa se identifican y se obtienen los datos que serán utilizados para entrenar el algoritmo de clasificación, para ello se ha tomado como punto de partida, realizar un análisis para entender el proceso que sigue normalmente un usuario para identificar aquellos productos que considera su competencia y así monitorearlos.

4.3.3 Limpieza de datos

Busca elevar la calidad de los datos al nivel requerido por las técnicas de análisis seleccionadas. Esto puede implicar la selección de los subconjuntos de datos limpios, la inserción de datos por defectos adecuados, o técnicas más ambiciosas tales como la estimación de datos faltantes mediante modelado (Microsoft, 2019).

4.3.4 Transformación de datos

Esta tarea incluye la construcción de operaciones de preparación de datos tales como la producción de atributos derivados o el ingreso de nuevos registros, o la transformación de valores para atributos existentes. Los atributos derivados son los atributos nuevos que son contruidos de uno o más atributos existentes en el mismo registro. Ejemplo: $\text{área} = \text{longitud} * \text{anchura}$ (Microsoft, 2019).

4.3.5 Integración de datos

Estos son los métodos por el cual la información es combinada de múltiples tablas o registros para crear nuevos registros o valores. La combinación de tablas se refiere a la unión simultánea de dos o más tablas que tienen información diferente sobre el mismo objeto. Ejemplo: una cadena de venta al público tiene una tabla con la información sobre las características generales de cada tienda (Por ejemplo, el espacio, el tipo de comercio), otra tabla con datos resumidos de las ventas (por ejemplo, el beneficio, el cambio porcentual en ventas desde el año anterior), y el otro con información sobre los datos demográficos del área circundante. Cada una de estas tablas contiene un registro para cada tienda. Estas tablas pueden ser combinadas simultáneamente en una nueva tabla con un registro para cada tienda, combinando campos de las tablas fuentes.

Los datos combinados también cubren agregaciones. La agregación se refiere a operaciones en la que nuevos valores son calculados de información resumida de múltiples registros y/o tablas. Por ejemplo, convirtiendo una tabla de compra de clientes donde hay un registro para cada compra en una tabla nueva donde hay un registro para cada cliente, con campos tales como el número de compras, el promedio de la cantidad de compra, el porcentaje de ordenes cobrados a tarjeta de crédito, el porcentaje de artículos bajo promoción, etc. (Microsoft, 2019).

4.3.6 Formatear datos

Formateando transformaciones se refiere a modificaciones principalmente sintácticas hechas a los datos que no cambian su significado, pero podría ser requerido por la herramienta de modelado. Algunas herramientas tienen requerimientos sobre el orden de los atributos, tales como el primer campo que es un único identificador para cada registro o el último campo es el campo resultado que el

modelo debe predecir. Podría ser importante cambiar el orden de los registros en el conjunto de datos. Quizás la herramienta de modelado requiere que los registros sean clasificados según el valor del atributo de resultado. Comúnmente, los registros del conjunto de datos son ordenados al principio de algún modo, pero el algoritmo que modela necesita que ellos estén en un orden moderadamente arbitrario. Por ejemplo, cuando se usa redes neuronales, esto es generalmente mejor para los registros para ser presentados en un orden aleatorio, aunque algunas herramientas manejen esto automáticamente sin la intervención explícita del usuario. Además, hay cambios puramente sintácticos hechos para satisfacer las exigencias de la herramienta de modelado específica. Ejemplos: el quitar de comas de adentro de campos de texto en ficheros de datos delimitados por coma, corta todos los valores a un máximo de 32 caracteres (Microsoft, 2019).

4.4 Definición del Modelo

Esta etapa consiste en generar y definir el modelo o los modelos de minería de datos utilizando los datos adquiridos en la etapa previa, para ello se aplica el algoritmo de clasificación Naive Bayes para detectar los registros que son considerados competencia sobre los datos previamente obtenidos y procesados durante las etapas previas. Se deben definir qué columnas de datos serán utilizadas por el modelo, un modelo de minería de datos simplemente es un contenedor que especifica las columnas que se usan para la entrada, el atributo que está prediciendo y parámetros que indican al algoritmo cómo procesar los datos.

El procesamiento de un modelo generalmente se denomina entrenamiento. El entrenamiento se refiere al proceso de aplicar algoritmos matemáticos específicos a estructuras de datos para extraer patrones. El patrón encontrado durante el proceso de entrenamiento dependerá de la elección de los datos de entrenamiento, el algoritmo seleccionado y la forma en que se configura el algoritmo.

Los parámetros se utilizan para optimizar cada algoritmo de modo que se puedan aplicar filtros a los datos de entrenamiento para utilizar un subconjunto de los datos y producir resultados diferentes. Una vez que los datos han pasado por el modelo, el objeto del modelo de minería de datos contiene los resúmenes y modelos que se pueden consultar o utilizar para la predicción.

El algoritmo de clasificación seleccionado para este proyecto es Naive Bayes, que, como se mencionó anteriormente en el marco teórico, es un poderoso algoritmo basado en la probabilidad y el teorema de Bayes de fácil implementación que permite la detección y clasificación de conjuntos de datos. basado en datos de entrenamiento u observaciones obtenidas previamente.

El lenguaje Python cuenta con diferentes librerías como Scikit Learn y Pandas las cuales simplifican el análisis y procesamiento de datos, así como también permiten utilizar los diferentes algoritmos para minería de datos entre ellos el que se plantea utilizar. La librería Scikit Learn considera lo siguiente para utilizar el algoritmo de Naive Bayes. Primero, los métodos Naive Bayes son un conjunto de algoritmos de aprendizaje supervisado que aplican el teorema de Bayes el cual asume una independencia condicional entre cada par de características dado el valor de la variable de clase. El teorema de Bayes establece la siguiente relación, dada la variable de clase y y el vector de característica dependiente x_1 mediante x_n .

$$p(C|F_1, \dots, F_n) = \frac{p(C) p(F_1, \dots, F_n|C)}{p(F_1, \dots, F_n)}.$$

Ecuación 1 Teorema de Bayes

En la práctica solo importa el numerador, ya que el denominador no depende de C y los valores de F_i son datos, por lo que el denominador es, en la práctica, constante. El numerador es equivalente a una probabilidad compuesta (Ecuación 2).

$$p(C, F_1, \dots, F_n)$$

Ecuación 2 Ejemplo de probabilidad compuesta

La ecuación 2 puede reescribirse como se muestra en la ecuación 3, aplicando repetidamente la definición de probabilidad condicional.

$$\begin{aligned} p(C, F_1, \dots, F_n) &= p(C) p(F_1, \dots, F_n|C) \\ &= p(C) p(F_1|C) p(F_2, \dots, F_n|C, F_1) \\ &= p(C) p(F_1|C) p(F_2|C, F_1) p(F_3, \dots, F_n|C, F_1, F_2) \\ &= p(C) p(F_1|C) p(F_2|C, F_1) p(F_3|C, F_1, F_2) p(F_4, \dots, F_n|C, F_1, F_2, F_3) \end{aligned}$$

Ecuación 3 Probabilidad compuesta aplicando probabilidad condicional

Utilizando el supuesto ingenuo de independencia condicional, se asume que cada F_i es independiente de cualquier F_j para $j \neq i$ cuando están condicionadas a C . Esto significa que la ecuación 3 puede representarse como la ecuación 4.

$$p(F_i|C, F_j) = p(F_i|C)$$

Ecuación 4 Aplicación de ingenuo de Naive Bayes

La probabilidad compuesta se expresa como puede observarse en la ecuación 5.

$$\begin{aligned} p(C, F_1, \dots, F_n) &= p(C) p(F_1|C) p(F_2|C) p(F_3|C) \dots \\ &= p(C) \prod_{i=1}^n p(F_i|C). \end{aligned}$$

Ecuación 5 Probabilidad compuesta

Aplicando los supuestos de la distribución condicional de C sobre las variables clasificatorias puede expresarse como se muestra en la ecuación 6, donde Z es un factor que depende solo de F_1, \dots, F_n , es decir, constante si los valores de F_i son conocidos.

$$p(C|F_1, \dots, F_n) = \frac{1}{Z} p(C) \prod_{i=1}^n p(F_i|C)$$

Ecuación 6 Naive Bayes expresado con distribución condicional

Aunque los clasificadores de Bayes ingenuos parecen demasiado simples, funcionan bien en muchas situaciones prácticas (como la clasificación de documentos y el filtrado de spam) y requieren una pequeña cantidad de datos de entrenamiento para estimar los parámetros necesarios

Los clasificadores Naive Bayes pueden ser extremadamente rápidos en comparación con métodos más sofisticados. El desacoplamiento de las distribuciones de características condicionales de clase significa que cada distribución se puede estimar de forma independiente como una distribución unidimensional. Esto, a su vez, ayuda a aliviar los problemas derivados de la maldición de la dimensionalidad. Por otro lado, aunque Naive Bayes se conoce como un clasificador decente, se sabe que es un mal estimador, por lo que las salidas de probabilidad no deben tomarse demasiado en serio (ScikitLearn, s.f.).

4.4.1 Validación del Modelo

La validación es el proceso de evaluar cuál sería el rendimiento de sus modelos de minería de datos con datos reales. Es importante que valide sus modelos de minería de datos entendiendo su calidad y sus características antes de implementarlos en un entorno de producción (msdn.microsoft.com, 2020).

La validación es el proceso de evaluar cuál sería el rendimiento de sus modelos de minería de datos con datos reales. Es importante que valide sus modelos de minería de datos entendiendo su calidad y sus características antes de implementarlos en un entorno de producción. Las medidas de minería de datos se suelen agrupar en las categorías de precisión, confiabilidad y utilidad (msdn.microsoft.com, 2020).

La precisión es una medida que indica hasta qué punto el modelo pone en correlación un resultado con los atributos de los datos que se han proporcionado. Existen varias medidas de precisión, pero todas ellas dependen de los datos que se utilicen. En realidad, podrían faltar valores o éstos ser aproximados, o incluso diferentes procesos podrían cambiar los datos. En particular, en la fase de exploración

y desarrollo, podría decidir aceptar una cierta cantidad de errores en los datos, sobre todo si éstos son suficientemente uniformes en sus características. Por ejemplo, un modelo que predice las ventas para un almacén determinado en base a las ventas pasadas puede estar muy correlacionado y ser muy preciso, incluso si ese almacén ha utilizado un método de contabilidad equivocado continuamente. Por tanto, es necesario equilibrar las mediciones de precisión mediante las valoraciones de confiabilidad (msdn.microsoft.com, 2020).

La confiabilidad evalúa la manera en la que se comporta un modelo de minería de datos en conjuntos de datos diferentes. Un modelo de minería de datos es confiable si genera el mismo tipo de predicciones o encuentra los mismos tipos generales de patrones independientemente de los datos de prueba que se proporcionen. Por ejemplo, el modelo que ha generado para el almacén que utilizó un método de contabilidad equivocado no podría extrapolarse correctamente a otros almacenes, y por tanto, no sería confiable (msdn.microsoft.com, 2020).

La utilidad incluye diferentes métricas que le indican si el modelo proporciona información útil. Por ejemplo, un modelo de minería de datos que pone en correlación la ubicación del almacén con las ventas podría ser preciso y fiable, pero podría no ser útil, ya que no se podría generalizar ese resultado si se agregaran más almacenes en la misma ubicación. Es más, no responde a la pregunta comercial fundamental de por qué ciertas ubicaciones tienen más ventas que otras. También podría descubrir que un modelo que parece correcto, en realidad no tiene sentido porque está basado en correlaciones cruzadas de los datos.

Existen varios enfoques para la validación de soluciones de minería de datos, que abarcan todas las fases de la metodología de prueba de la minería de datos.

- Crear particiones de los datos de los conjuntos de prueba y entrenamiento.
- Filtrar modelos para entrenar y probar combinaciones diferentes de los mismos datos de origen.
- Medir la mejora respecto al modelo predictivo y la ganancia. Un gráfico de mejora respecto al modelo predictivo es un método para visualizar la mejora que obtendrá de usar un modelo de minería de datos, si lo compara con una estimación aleatoria.
- Realizar una validación cruzada de los conjuntos de datos
- Generar matrices de clasificación. Estos gráficos ordenan las estimaciones buenas y malas en una tabla, lo que permite analizar rápida y fácilmente con qué precisión predice el modelo el valor de destino.
- Crear gráficos de dispersión para evaluar el ajuste de una fórmula de regresión.
- Crear gráficos de beneficios que permiten asociar ganancias o costos financieros con el uso de cierto modelo de minería de datos, para poder evaluar el valor de las recomendaciones.

Estas métricas no pretenden responder a la pregunta de si el modelo de minería de datos resuelve sus preguntas empresariales, sino que proporcionan medidas objetivas que puede usar para evaluar la confiabilidad de los datos para los análisis predictivos, y le ofrecen ayuda a la hora de decidir si debe usar una iteración determinada en el proceso de desarrollo (msdn.microsoft.com, 2020).

4.5 Detección Automática

En esta etapa se utiliza el modelo entrenado con el algoritmo de Naive Bayes para analizar nuevas tuplas agregadas al conjunto de datos, por ejemplo, considerando que los valores de los atributos de una nueva tupla son:

- (“1”, “2”, “1”, “2”, “1”)

Considerando el orden de las columnas mostradas en el Cuadro 1, el algoritmo debe de calcular la probabilidad de que este pertenezca a la clase Competencia, para ello se considera lo siguiente, existen 2 clases identificadas para la etiqueta de clase Competencia, $C = \text{si}$ y $C = \text{no}$. Sustituyendo en la ecuación (2) para cada uno de ellas se obtienen las ecuaciones (4) y (5).

- $P(C=\text{si} | T) \sim P(T | C=\text{si}) * P(C=\text{si})$ (4)
- $P(C=\text{no} | T) \sim P(T | C=\text{no}) * P(C=\text{no})$ (5)

En el primer paso se calcula $P(C=\text{si})$ y $P(C=\text{no})$ utilizados en las ecuaciones (4) y (5).

Para este ejemplo, se están considerando 17 registros correctamente identificados.

- $P(C=\text{si}) = 10/17 = 0.59$
- $P(C=\text{no}) = 7/17 = 0.41$

En el segundo paso calculamos la probabilidad de $P(T | C=\text{si})$ y $P(T | C=\text{no})$ al sustituir los valores individuales de probabilidad por cada valor del atributo en la ecuación (3).

- $P(R=1|C=\text{si}) = 4/10=0.4$
- $P(P=2|C=\text{si}) = 2/10=0.2$
- $P(B=1|C=\text{si}) = 2/10=0.2$
- $P(E=2|C=\text{si}) = 3/10=0.3$
- $P(T=1|C=\text{si}) = 1/10=0.1$
- $P(T | C=\text{si}) = 0.4*0.2*0.2*0.3*0.1=0.00048$

- $P(R=1|C=no) = 3/7=0.429$
- $P(P=2|C=no) = 7/7=1$
- $P(B=1|C=no) = 5/7=0.714$
- $P(E=2|C=no) = 4/7=0.571$
- $P(T=1|C=no) = 6/7=0.857$
- $P(T | C=no) = 0.429*1*0.714*0.571*0.857=0.14988$

Como último paso se realiza la multiplicación representada por las ecuaciones (4) y (5) para obtener el resultado final. Dado que $P(C=no | T) > P(C=sí | T)$ la clase de la nueva fila es identificada como $C = no$.

- $P(C=sí | T) \sim P(T | C=sí) * P(C=sí) = 0.0005*0.59=0.000282$
- $P(C=no | T) \sim P(T | C=no) * P(C=no) = 0.1499*0.41=0.0616$

Capítulo 5. Desarrollo e Implementación

5.1 Desarrollo de la aplicación

El seguimiento de competencia por el proyecto MLCOMPET se encarga de obtener y comparar registros de publicaciones en MercadoLibre y comparar los productos de otros vendedores con los del usuario para detectar posibles competidores y realizar acciones que le permitan al usuario posicionar mejor sus productos.

En el proyecto MLCOMPET se desarrolla como una aplicación web que integra librerías de machine learning que le permite hacer un análisis eficiente de la competencia. Al ser una aplicación web cuenta con ciertas características como:

- Acceso desde cualquier dispositivo conectado a internet
- Información siempre disponible
- Análisis y Actualización de datos en tiempo real

Para el desarrollo de este proyecto se tomó la decisión de utilizar el lenguaje Python por ser un lenguaje bastante robusto y con una programación multiparadigma la cual permite trabajarlo como un proyecto web y que además es compatible con las librerías de Machine Learning utilizadas en el proyecto.

Adicionalmente se ha elegido utilizar un framework de desarrollo con el fin de ser mas eficientes a la hora de realizar el proyecto. El framework utilizado es Django en su versión 2.0 por su madurez y amplia aceptación por la comunidad de desarrolladores, a continuación, se puede observar a continuación los diferentes lenguajes y librerías utilizadas durante el desarrollo (Tabla 2).

Tipo	Nombre	Versión
Lenguaje	Python	3.7
Framework	Django	2.0
Base de Datos	Sqlite	3
Librería	Scikit-Learn	0.6
Librería	MercadoLibre SDK	1
Librería Front End	Bootstrap	4.0

Tabla 2 Listado de dependencias

5.1.1. Instalación de Python

Al ser un proyecto desarrollado en Python, es necesario instalar un ambiente de desarrollo con las librerías y módulos necesarios para poder realizarlo de una forma eficiente. El primer paso es descargar e instalar el lenguaje de desarrollo, en este caso Python, para ello hay que abrir el sitio web oficial <https://www.python.org/> y en la sección de descargas seleccionar la versión adecuada para el sistema operativo en el que se va a trabajar (Figura 28).



Figura 28 Python para Windows

Para el desarrollo de este proyecto se ha elegido utilizar Windows 10, por lo que se selecciona la versión más reciente disponible y se descarga (Figura 29).



Figura 29 Descarga de Python

Una vez descargado el software se debe de ejecutar el programa y seguir las instrucciones para su instalación (Figura 30).



Figura 30 Instalación de Python

Una parte importante al instalar Python en el equipo de desarrollo es incluir el directorio de instalación al PATH o variable de entorno del sistema, para ello se debe seleccionar la opción “Add Python 3.9 to PATH” (Figura 31).



Figura 31 Agregar Python al PATH

Una vez concluida la instalación se puede verificar en una terminal del sistema, escribiendo la palabra python o python3 seguido por -V, si todo esta en orden se mostrara la versión de python instalada (Figura 32).



```
cmd
C:\Users\glvega
λ python -V
Python 3.8.5
```

Figura 32 Verificar instalación de Python

5.1.2 Creación de Entorno Virtual para el proyecto

Una vez instalado Python en el equipo de desarrollo lo siguiente es generar un Entorno Virtual para el proyecto (*Virtual Environment*). Los entornos virtuales permiten encapsular las dependencias de un proyecto y mantenerlas aisladas del resto de proyectos o aplicaciones que coexisten en el mismo equipo.

Para crear el entorno virtual basta con ejecutar el siguiente comando en la terminal, el cual internamente va a hacer una copia de la instalación de python que será dedicada para el proyecto sin afectar a las demás aplicaciones como se había mencionado con anterioridad (Figura 33).



```
C:\Users\glvega
λ python -m venv env
```

Figura 33 Creación del entorno virtual para el proyecto.

Una vez creado el entorno virtual se debió haber generado una carpeta en nuestro equipo, con una estructura de archivos similar a la que se muestra a continuación (Figura 34).

Name	Date modified	Type
Include	11/24/2020 2:46 PM	File folder
Lib	11/24/2020 2:46 PM	File folder
Scripts	11/24/2020 2:46 PM	File folder
pyvenv.cfg	11/24/2020 2:46 PM	CFG File

Figura 34 Estructura de archivos de un entorno virtual de Python

Para comenzar a utilizar el entorno virtual este debe de activarse, para ello es necesario ejecutar el comando “activate” en la consola del sistema (Figura 35).

```
C:\Users\glvega
λ cd env\

C:\Users\glvega\env
λ Scripts\activate.bat |
```

Figura 35 Activar entorno virtual de Python

Una vez activado el entorno virtual, el apuntador en la terminal va cambiar mostrando ahora el nombre del entorno creado, para este proyecto se llama ‘env’ (Figura 36).

```
C:\Users\glvega\env
(env) λ |
```

Figura 36 Entorno virtual activo

5.1.3 Instalación de Django

El siguiente paso es instalar las librerías y el framework necesarios para comenzar el desarrollo, para ello se hace uso del gestor de paquetes de python llamado pip, el cual se utiliza para instalar o desinstalar librerías del ambiente de desarrollo que se esté utilizando, para este proyecto, las librerías se instalaran en el entorno virtual activo.

La primera librería que se debe instalar es Django, la cual a su vez instalara algunas librerías adicionales de las cuales depende. Para instalar cualquier librería con pip se debe ejecutar el comando “*pip install nombre_libreria*”. Para instalar Django quedaría de la siguiente forma (Figura 37).

```
C:\Users\glvega\env
(env) λ pip install django
```

Figura 37 Instalar Django con el gestor de paquete pip

Al ejecutar el comando anterior comenzara la descarga e instalación de Django y sus dependencias, en caso de existir algún problema con la instalación, se debe mostrar un mensaje indicando el problema (Figura 38).

```
C:\Users\glvega
(env) λ pip install django
Collecting django
  Downloading Django-3.1.3-py3-none-any.whl (7.8 MB)
  |-----| 7.8 MB 409 kB/s
Collecting asgiref<4,>=3.2.10
  Downloading asgiref-3.3.1-py3-none-any.whl (19 kB)
Collecting sqlparse>=0.2.2
  Downloading sqlparse-0.4.1-py3-none-any.whl (42 kB)
  |-----| 42 kB 1.7 MB/s
Collecting pytz
  Downloading pytz-2020.4-py2.py3-none-any.whl (509 kB)
  |-----| 509 kB 504 kB/s
Installing collected packages: asgiref, sqlparse, pytz, django
Successfully installed asgiref-3.3.1 django-3.1.3 pytz-2020.4 sqlparse-0.4.1
WARNING: You are using pip version 20.1.1; however, version 20.2.4 is available.
You should consider upgrading via the 'c:\users\glvega\appdata\local\programs\python\pyth
ll --upgrade pip' command.
```

Figura 38 Descarga e Instalación de Django

Para verificar las librerías instaladas en el ambiente de desarrollo se puede ejecutar el comando “*pip freeze*” el cual mostrara un listado con el nombre y la versión de los paquetes instalados (Figura 39).

```
C:\Users\glvega
(env) λ pip freeze
asgiref==3.3.1
astroid==2.4.2
colorama==0.4.3
Django==3.1.3
isort==5.5.4
lazy-object-proxy==1.4.3
mccabe==0.6.1
pylint==2.6.0
pytz==2020.4
six==1.15.0
sqlparse==0.4.1
toml==0.10.1
wrapt==1.12.1
```

Figura 39 Librerías instaladas ejecutando el comando pip freeze

El comando pip freeze también se utiliza para generar archivos de requerimientos para entornos virtuales de modo que cuando el proyecto pase a producción se puedan instalar las mismas librerías y versiones instaladas en desarrollo, para ello se modifica un poco el comando anterior y se agrega el nombre de un archivo de texto en donde se guardarán esos resultados (Figura 40).

```
C:\Users\glvega
(env) λ pip freeze > requirements.txt
```

Figura 40 Generación de archivo de dependencias

El archivo generado debe agregarse a la carpeta del proyecto, en este caso se genera en la ruta donde apunta la terminal al momento de ejecutarse (Figura 41).

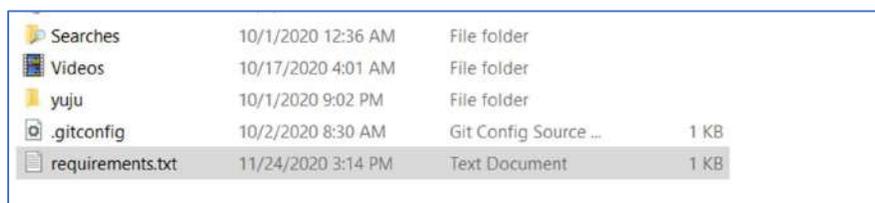


Figura 41 Ubicación del archivo de dependencias

5.1.4 Instalación de otras librerías

Para el desarrollo de este proyecto fue necesario utilizar algunas otras librerías, las cuales fueron necesarias para la correcta operación del sistema, entre las cuales se encuentran *gunicorn* y *psycopg2*, las cuales se encargan de la gestión del servicio web y la base de datos respectivamente, el listado completo de las dependencias del proyecto se muestra en el cuadro siguiente.

asgiref==3.2.10	certifi==2018.8.24
chardet==3.0.4	dj-database-url==0.5.0
Django==3.1.2	django-common-helpers==0.9.2
django-cron==0.5.1	django-heroku==0.3.1
gunicorn==19.9.0	idna==2.7
psycopg2==2.7.5	pytz==2018.5
requests==2.19.1	sqlparse==0.4.1
urllib3==1.23	whitenoise==4.1

Tabla 3 Dependencias del proyecto

5.1.5 Repositorio del proyecto en Github

Con la finalidad de tener una gestión más eficiente del proyecto y de los cambios en el código, se ha decidido utilizar un sistema de control de versiones, el cual permite llevar un registro cronológico de los cambios realizados al proyecto.

Además, al utilizar un sistema de control de versiones se facilita el proceso de implementación ya que el proyecto se puede lanzar a cualquier servidor en línea configurando el acceso al repositorio, por lo que la liberación de cambios a producción es más eficiente.

GitHub es una plataforma para alojar proyectos que utilizan el sistema de control de versiones de Git. Se utiliza principalmente para crear el código fuente de

un programa informático. El software que ejecuta GitHub está escrito en Ruby on Rails. Desde enero de 2010, GitHub opera con el nombre de GitHub, Inc. (Figura 42).



Figura 42 Github plataforma para control de versiones

Para poder utilizar Github, es necesario contar con una cuenta, por lo que es necesario registrarse (Figura 43).

Figura 43 Formulario de acceso a Github

Una vez que se tiene una cuenta de Github, se puede acceder a la plataforma y crear un repositorio para su código.

El proyecto se encuentra registrado por el nombre de mlcompet bajo la cuenta del usuario, para acceder al repositorio se puede buscar por el nombre del proyecto o por medio de la siguiente URL <https://github.com/glopzvega/mlcompet> (Figura 44).

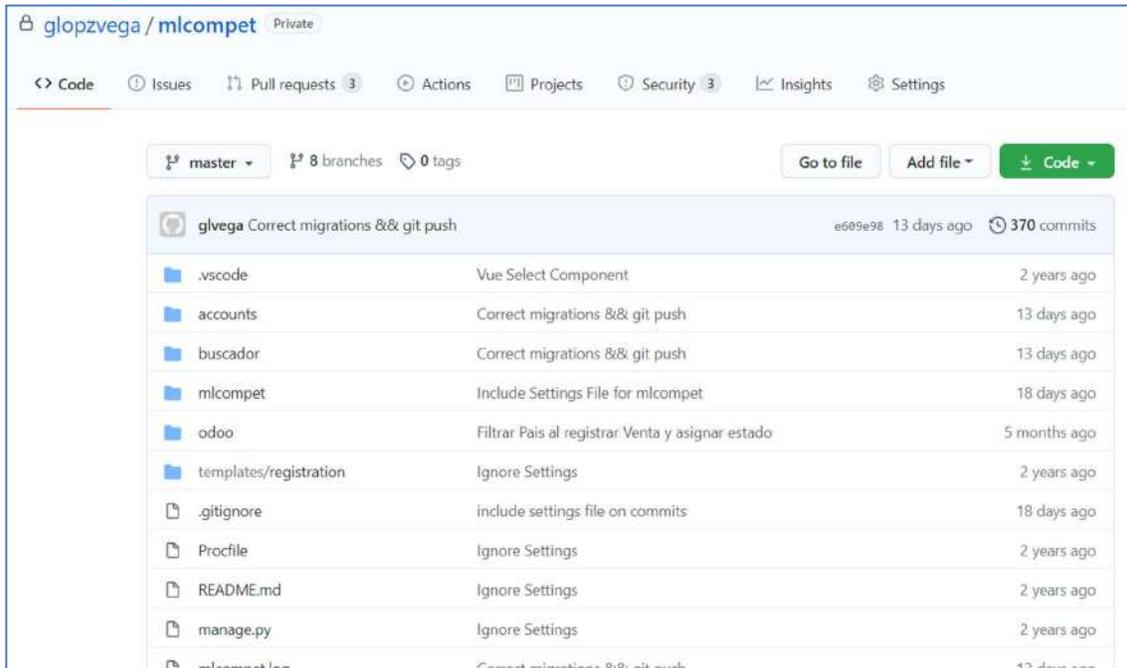


Figura 44 Repositorio de Github para el proyecto.

En el repositorio se pueden consultar los cambios y versiones que ha tenido el proyecto desde su creación, para ello podemos utilizar los controles que aparecen en la parte superior del repositorio (Figura 45).

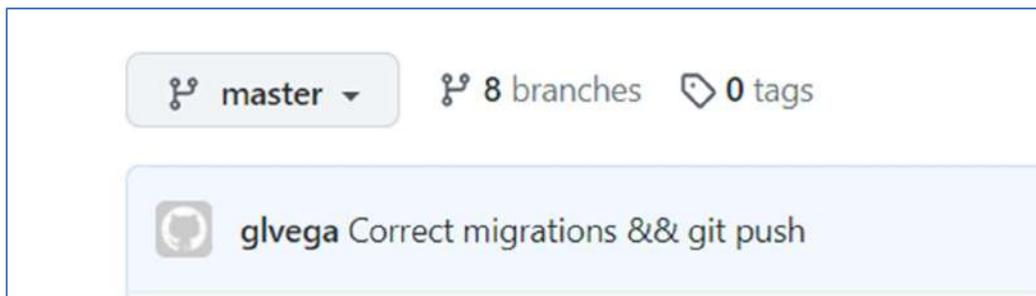


Figura 45 Controles del repositorio

5.1.6 Creación del proyecto con Django

El siguiente paso es crear el proyecto con ayuda de Django, para ello se ejecuta el comando `django-admin.exe startproject nombre_proyecto` en la terminal del sistema, el comando debe ejecutarse en la ubicación donde se desee crear el proyecto (Figura 46).

```
C:\Users\glvega
(env) λ django-admin.exe startproject mlcompet2
```

Figura 46 Creación del proyecto con Django.

Una vez creado el proyecto se habrá generado un directorio con el nombre del proyecto indicado al momento de ejecutar el comando, el directorio tendrá una estructura de archivos similar a la mostrada en la siguiente imagen (Figura 47).

Name	Date modified	Type	Size
mlcompet2	11/24/2020 3:53 PM	File folder	
manage.py	11/24/2020 3:53 PM	Python Source File	1 KB

Figura 47 Proyecto creado con Django

El documento con el nombre `manage.py` es un archivo que se ejecuta para realizar múltiples acciones con proyectos desarrollados con Django, entre ellas se encuentran, la creación de usuarios, creación de módulos nuevos, aplicación de cambios a la base de datos, entre otros. Por otro lado, la carpeta que aparece con el nombre del proyecto (`mlcompet`), incluye principalmente archivos para la configuración del proyecto (Figura 48), el archivo de configuración se revisa a mayor detalle en los apartados siguientes.

Name	Date modified	Type	Size
init.py	11/24/2020 3:53 PM	Python Source File	0 KB
asgi.py	11/24/2020 3:53 PM	Python Source File	1 KB
settings.py	11/24/2020 3:53 PM	Python Source File	4 KB
urls.py	11/24/2020 3:53 PM	Python Source File	1 KB
wsgi.py	11/24/2020 3:53 PM	Python Source File	1 KB

Figura 48 Archivos de configuración del sistema

5.1.7 Inicialización del proyecto

El siguiente paso en el desarrollo es inicializar el proyecto, para ello hay que entrar a la carpeta del proyecto y ejecutar el comando `python manage.py migrate` el cual va a crear la base de datos del proyecto (Figura 49).

```
C:\Users\glvega\mlcompet2
λ python manage.py migrate
Operations to perform:
  Apply all migrations: admin, auth, contenttypes, sessions
Running migrations:
  Applying contenttypes.0001_initial... OK
  Applying auth.0001_initial... OK
  Applying admin.0001_initial... OK
  Applying admin.0002_logentry_remove_auto_add... OK
  Applying admin.0003_logentry_add_action_flag_choices... OK
  Applying contenttypes.0002_remove_content_type_name... OK
  Applying auth.0002_alter_permission_name_max_length... OK
  Applying auth.0003_alter_user_email_max_length... OK
  Applying auth.0004_alter_user_username_opts... OK
  Applying auth.0005_alter_user_last_login_null... OK
  Applying auth.0006_require_contenttypes_0002... OK
  Applying auth.0007_alter_validators_add_error_messages... OK
  Applying auth.0008_alter_user_username_max_length... OK
  Applying auth.0009_alter_user_last_name_max_length... OK
  Applying auth.0010_alter_group_name_max_length... OK
  Applying auth.0011_update_proxy_permissions... OK
  Applying auth.0012_alter_user_first_name_max_length... OK
  Applying sessions.0001_initial... OK
```

Figura 49 Inicialización del proyecto con el comando `migrate`

Como se mencionó anteriormente, el comando `migrate` genera la estructura por default de la base de datos para un proyecto de Django, esta base de datos inicial incluye tablas para el manejo de usuarios y permisos en la aplicación, entre algunas otras para el manejo de las vistas y módulos instalados.

5.1.8 Levantar el servidor web del proyecto

Una vez inicializado el proyecto, el siguiente paso es levantar el servidor web para poder acceder al proyecto desde el navegador, para ello, se debe ejecutar el comando `python manage.py runserver` desde la terminal del sistema (Figura 50).

```
C:\Users\glvega\mlcompet2
λ python manage.py runserver
Watching for file changes with StatReloader
Performing system checks...

System check identified no issues (0 silenced).
November 24, 2020 - 17:02:20
Django version 3.1.3, using settings 'mlcompet2.settings'
Starting development server at http://127.0.0.1:8000/
Quit the server with CTRL-BREAK.
```

Figura 50 Levantar servidor web del proyecto

Por default el servidor web se inicia en el puerto 8000, sin embargo, este puerto puede personalizarse para que se utilice el que sea más conveniente para el usuario, por ejemplo, para este proyecto se ha utilizado el puerto 80 el cual es el default para aplicaciones web. Para configurar el puerto se debe enviar al momento de ejecutar el comando (Figura 51).

```
C:\Users\glvega\mlcompet2
λ python manage.py runserver 80
Watching for file changes with Sta
```

Figura 51 Levantar servidor web en puerto 80

Aunque el comando `runserver` permite levantar un servidor web que sirve la aplicación desarrollada a través del navegador, cabe mencionar que este comando solo se utiliza en desarrollo, ya que por sus características no es suficiente para ambientes productivos donde se necesita una mejor gestión de accesos y recursos, para ello se utiliza la librería `gunicorn` de la cual se hablara más adelante.

5.1.9 Acceder al proyecto desde navegador web

Una vez levantado el servidor web del proyecto, es posible acceder a la aplicación desde un navegador web, por ejemplo Google Chrome o Mozilla Firefox, para abrir la aplicación se debe introducir en la barra de navegación la siguiente URL <http://localhost/> o <http://127.0.0.1/>, la cual muestra la aplicación por defecto del proyecto de Django que se ha creado (Figura 52).

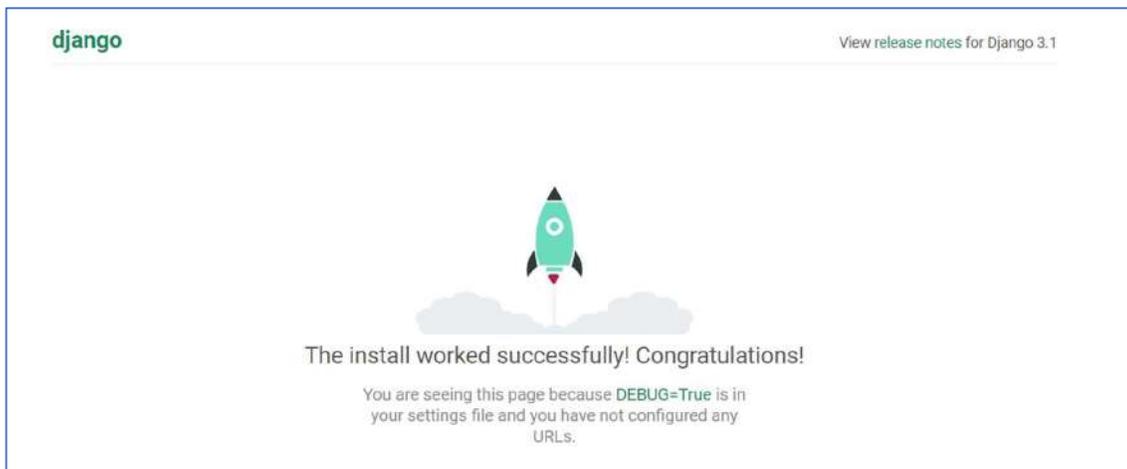


Figura 52 Aplicación corriendo desde el navegador web

5.1.10 Acceder al panel administrativo del proyecto

Django cuenta con una poderosa herramienta que lo diferencia de otros frameworks, la cual es su panel administrativo, donde se pueden gestionar diferentes elementos como las tablas, usuarios, permisos o simplemente generar nuevos registros o realizar búsquedas. Para poder acceder al *Admin* del proyecto es necesario contar con un usuario con privilegios de administrador, para ello se debe ejecutar el comando `python manage.py createsuperuser` en la terminal del sistema (Figura 53).

```
C:\Users\glvega\mlcompet2
λ python manage.py createsuperuser
Username (leave blank to use 'glvega'):
Email address:
Password:
Password (again):
Error: Blank passwords aren't allowed.
Password:
Password (again):
Superuser created successfully.
```

Figura 53 Creación de super usuario del sistema.

Una vez ejecutado el comando, el sistema solicitará llenar alguna información como el nombre de usuario, email y password. Una vez completado, se notificará que el usuario se ha creado satisfactoriamente. Habiendo creado el usuario ya se puede acceder al Admin del sistema, para ello hay que acceder a la URL <http://localhost/admin/> en donde se solicitarán los datos del super usuario registrado anteriormente (Figura 54).

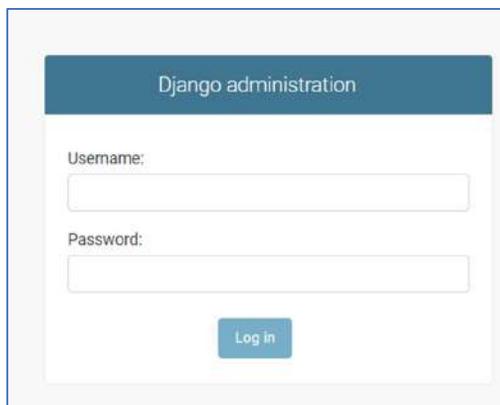


Figura 54 Formulario de Acceso al Admin del sistema

Una vez que se ha logueado al sistema se pueden observar los modelos registrados actualmente en la base de datos, los cuales por default son Usuarios y Grupos con los cuales se gestiona el acceso a la aplicación (Figura 55).

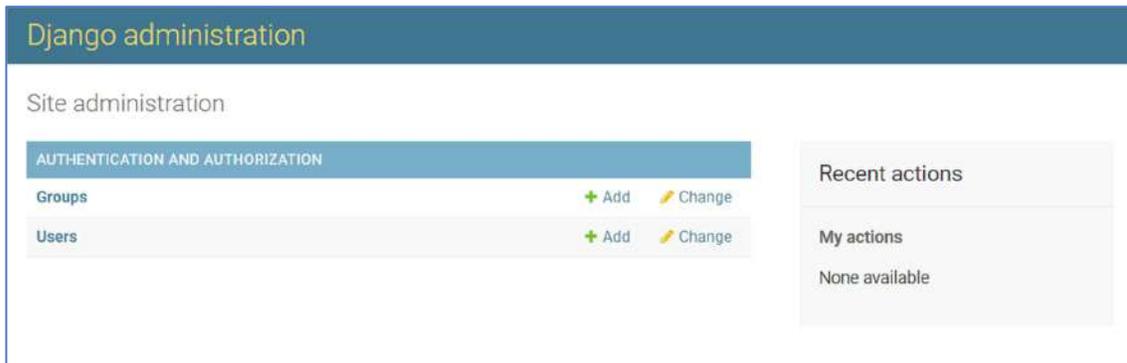


Figura 55 Panel de control del Admin del sistema

La vista del Admin o panel de control del sistema se puede personalizar conforme se agregan funcionalidades y características al proyecto, en los apartados siguientes se revisa más a detalle como personalizar el Admin del proyecto.

5.1.11 Configuración del Proyecto

Para cada proyecto desarrollado en Django Framework, se tiene un archivo de configuración, el cual permite definir las características generales acerca del funcionamiento de la aplicación. Entre las configuraciones que se pueden definir se tiene la conexión a la base de datos, rutas a los diferentes recursos como archivos estáticos, plantillas e imágenes, módulos instalados, acceso a la aplicación desde dominios específicos, datos de conexión a Mercado Libre, entre otros (Figura 56).

```
# Application definition

INSTALLED_APPS = [
    'buscador',
    'accounts',
    'django_cron',
    'django.contrib.admin',
    'django.contrib.auth',
    'django.contrib.contenttypes',
    'django.contrib.sessions',
    'django.contrib.messages',
    'django.contrib.staticfiles',
]
```

Figura 56 Configuraciones del proyecto

El archivo de configuración se encuentra en el directorio que tiene el mismo nombre del proyecto se creó, y tiene el nombre de settings.py (Figura 57).

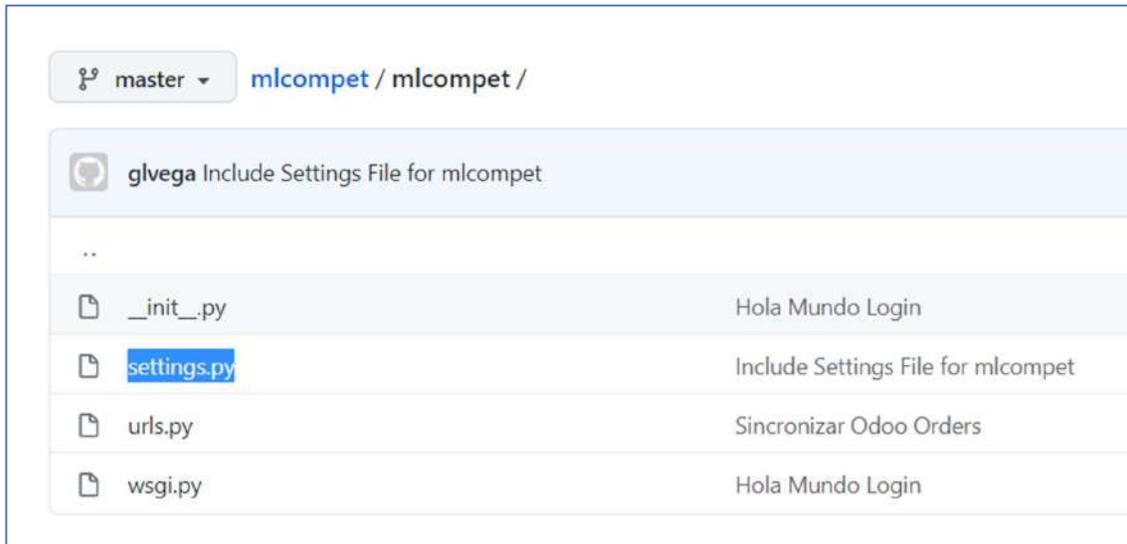


Figura 57 Archivo de configuración del proyecto

5.2 Integración con Mercado Libre

El flujo de conexión explicado anteriormente es el más adecuado para las aplicaciones desarrolladas en lenguajes como Java o Python, en donde se recomienda utilizar las librerías SDK desarrolladas por Mercado Libre para facilitar el desarrollo y reducir la complejidad en el flujo de autorización ya que se utiliza el protocolo OAuth (Figura 58).

The screenshot shows a table titled 'Librerías recomendadas' with three rows of recommended libraries:

Librería	Apps creadas	Requests	Acción
php PHP	6504	61.8B	Descargar
Python	514	51.8B	Descargar
.Net	723	21.8B	Descargar

Figura 58 Librerías disponibles para integrar con Mercado Libre

Para el desarrollo de esta aplicación se ha elegido la librería para Python ya que este lenguaje cuenta con las funciones necesarias para aplicar los diferentes algoritmos de inteligencia artificial y análisis de datos que se utilizan más adelante en el desarrollo de este proyecto.

La librería Python-SDK para Mercado Libre se encuentra alojada en el repositorio de código de la plataforma GitHub como puede observarse a continuación (Figura 59).

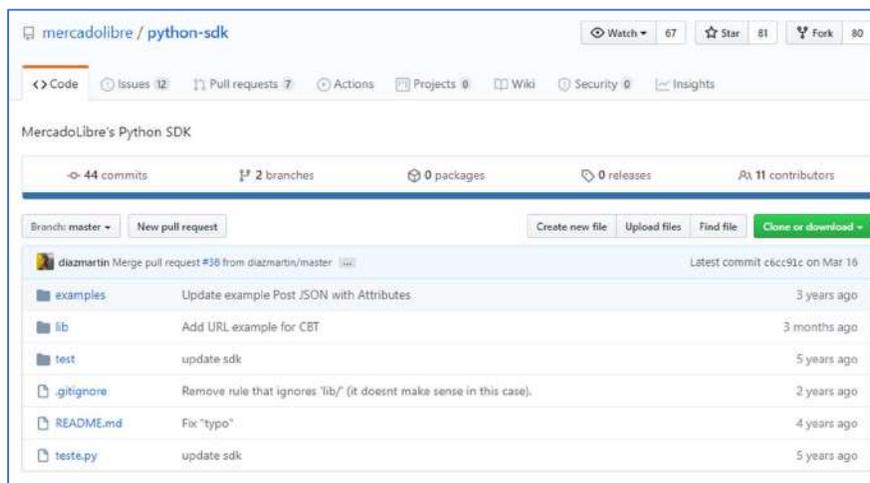


Figura 59 Repositorio de Código en GitHub de Mercado Libre

5.2.1 Módulo de Mis Productos

El módulo de **Mis Productos** cuenta con un campo de búsqueda, un botón para obtener los productos y una tabla donde se muestran los resultados de los productos obtenidos (Figura 60). El campo de búsqueda permite consultar los datos de productos que se encuentren actualmente en la base de datos o buscarlos directamente a Mercado Libre. Si se desea obtener todo el catálogo se presiona el botón de Sincronizar.



Figura 60 Campo de Búsqueda Modulo de Mis Productos

Para el campo de búsqueda se han utilizado los métodos de consultar productos por usuario, en donde se proporciona el USER_ID correspondiente al usuario, el USER_ID es el identificador único del usuario y se obtiene al registrar la aplicación en la plataforma de Mercado Libre (Figura 61).



Figura 61 Método para consultar productos por vendedor

El método anterior cuenta con la limitante de que funciona solamente cuando el usuario cuenta con menos de 1000 productos publicados, en caso contrario deberá emplearse otro método el cual ha sido utilizado en este módulo ya que, para este proyecto, el usuario en cuestión tiene más de 10,000 productos publicados en la plataforma por lo que debe agregarse un parámetro adicional en la consulta llamado SEARCH_TYPE con el valor de SCAN (Figura 62).



Figura 62 Método de búsqueda de productos mayor a 1000 registros.

En ambos casos la consulta debe ir acompañada del parámetro ACCESS_TOKEN el cual ha sido obtenido previamente (Figura 63).

```

def _get_user_items_meli(request):
    clienteML = False
    params = {"search_type" : "scan"}

    clientes = ClientML.objects.all()
    if clientes:
        clienteML = clientes[0]
        if clienteML.scroll_id:
            params.update({"scroll_id" : clienteML.scroll_id})

    url = '/users/%s/items/search' % '1010101010'

```

Figura 63 Ejemplo de consulta al API REST para el método de búsqueda

Una vez realizada la consulta a Mercado Libre a través de su API REST está devuelve una respuesta en formato JSON (Figura 64).

```

{
  "start_time": "2017-09-12T20:59:10.000Z",
  "seller_id": 3218374,
  "accepts_mercadopago": true,
  "attributes": [
    {
      "attribute_group_name": "Otros",
      "value_name": "Dun"
    }
  ],
  "automatic_relist": false,
  "available_quantity": 1,
  "base_price": 203,
  "buying_mode": "buy_it_now",
  "catalog_product_id": null,
  "category_id": "MLM45058",
  "condition": "new",
  "coverage_areas": [],
  "currency_id": "MXN",
  "date_created": "2017-09-12T20:59:10.000Z",
  "date_last_update": "2019-06-12 01:16"
}

```

Figura 64 Datos obtenidos de Mercado Libre en formato JSON

La respuesta obtenida se procesa y se almacena en la base de datos con ayuda del framework de Django (Figura 65).

```

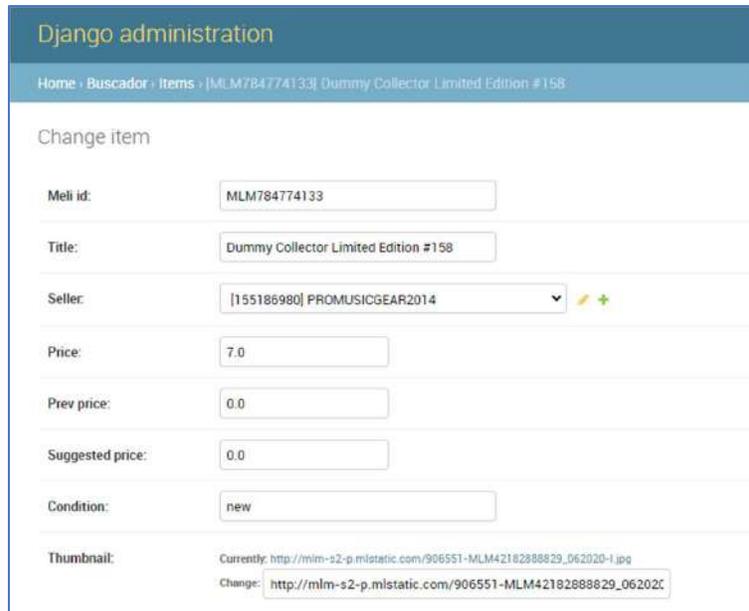
url = '/users/%s/items/search' % '155186980'
response = b_views._meli_call(url, params=params)

if response:
    for item in response["results"]:
        item = b_views._get_item_meli(item, True)
        newItem = b_views.createItem(item, "", True)

```

Figura 65 Ejemplo de procesamiento de la respuesta obtenida y almacenada

El modelo de datos para almacenar los productos del usuario se nombró ITEMS e incluye los siguientes atributos: id, nombre, vendedor, precio, condición, imagen, categoría, tipo de publicación, estatus, entre algunos otros datos que son útiles para este proyecto. Los datos pueden observarse con ayuda del panel administrativo de Django (Figura 66).



The screenshot displays the Django administration interface for editing a product. The page title is 'Django administration' and the breadcrumb trail is 'Home > Buscador > Items > [MLM784774133] Dummy Collector Limited Edition #158'. The form is titled 'Change item' and contains the following fields:

- Meli id:** MLM784774133
- Title:** Dummy Collector Limited Edition #158
- Seller:** [155186980] PROMUSICGEAR2014
- Price:** 7.0
- Prev price:** 0.0
- Suggested price:** 0.0
- Condition:** new
- Thumbnail:** Currently: http://mlm-s2-p.mlstatic.com/906551-MLM4218288829_062020-l.jpg
Change: http://mlm-s2-p.mlstatic.com/906551-MLM4218288829_062020-l.jpg

Figura 66 Datos del producto vistos con panel administrativo de Django.

Una vez que los registros obtenidos han sido almacenados en la base de datos, estos se visualizan en la tabla del módulo de **Mis Productos** la cual muestra información de cada uno de ellos. Los registros se van visualizando por medio de paginación, cada página muestra 50 registros. El total de productos se muestra en la parte superior de la tabla (Figura 67).

Total : 50 de 11250, Página: 1 de 226

Imagen	Producto	Precio	Estatus	Odoo
	1 Parche Power Beat Blanco Liso Tamaño: 12 Uk-0212-ba-10p ✓ MLM759593376 ✓ PMG2180_12 AG FREE Mercado	\$859.00	ACTIVO DROPS ESPECIAL	Editar 67410
	1 Parche Power Beat Blanco Liso Tamaño: 13 Uk-0213-ba-10p ✓ MLM759598944 ✓ PMG2180_13 AG FREE Mercado	\$909.00	ACTIVO DROPS ESPECIAL	Editar 67411
	1 Parche Power Beat Blanco Liso Tamaño: 14 Uk-0214-ba-10p ✓ MLM759598990 ✓ PMG2180_14 AG FREE Mercado	\$999.00	ACTIVO DROPS ESPECIAL	Editar 67412
	1 Parche Power Beat Blanco Liso Tamaño: 15 Uk-0215-ba-10p ✓ MLM759593422 ✓ PMG2180_15 AG FREE Mercado	\$999.00	ACTIVO DROPS ESPECIAL	Editar 67413

Figura 67 Tabla de Productos con Paginación

5.2.2 Módulo Mis Búsquedas

Una vez que los productos del usuario han sido obtenidos y se encuentran almacenados en la base de datos, el siguiente paso es obtener los productos de los competidores, para ello se desarrolló el módulo de **Mis Búsquedas** el cual permite al usuario hacer consultas a Mercado Libre con palabras clave (pueden ser nombre del producto, marca, modelo, etc.) con la cual se obtienen todas las publicaciones disponibles relacionadas a la búsqueda realizada (Figura 68).

Buscar Productos:

Envío: Todos

Exposición: Todos

Figura 68 Módulo de Mis Búsquedas

El módulo incluye filtros que permiten al usuario delimitar la búsqueda por medio de su tipo de envío y el nivel de exposición dentro de la plataforma, con lo que la búsqueda es más precisa. El resultado de esta búsqueda se muestra en forma de

listado en la parte inferior de módulo e incluye datos como el nombre del producto, el precio y el vendedor, así también como la cantidad disponible y la cantidad vendida, datos que son utilizados más adelante (Figura 69).

Producto	Precio
<input type="checkbox"/>  Guitarra Electrica Ibanez Grx55b Negra FAMUSICCLUB MLM768377775 Envío Gratis! Mercado de Pago	Precio: 4399 Tipo Listado: GOLD SPECIAL Disponibles: 1 Vendidos: 2 Envío: DROP OFF
<input type="checkbox"/>  Guitarra Electrica Ibanez Ps40 Paul Stanley LAFLAUTAMUSICAL MLM768769526 Envío Gratis! Mercado de Pago	Precio: 7500 Tipo Listado: GOLD PRO Disponibles: 1 Vendidos: 0 Envío: DROP OFF
<input type="checkbox"/>  Guitarra Electrica Ibanez Grx55b-ye Gio Series Amarilla POLVORA MUSIC MLM783533852 Envío Gratis! Mercado de Pago	Precio: 5126 Tipo Listado: GOLD PRO Disponibles: 1 Vendidos: 0

Figura 69 Resultado de búsqueda del usuario

En este listado el usuario también puede observar que productos de su catálogo coinciden con la búsqueda realizado, además de la información de las diferentes publicaciones en forma general. Con esta información el usuario puede comparar su producto con el resto de productos disponibles y determinar en base a su experiencia cuál es su competencia (Figura 70).

Producto	Precio
<input type="checkbox"/>  Guitarra Acustica Curva Nuevos Accesorios Completos Gratis GUITARRAS_PARACHO_MICH MLM620736719 Envío Gratis! Mercado de Pago	Precio: 1049 Tipo Listado: GOLD PRO Disponibles: 500 Vendidos: 500 Envío: DROP OFF

Figura 70 Vista general de los datos de productos

En esta etapa del proyecto el usuario selecciona aquellos productos que considera como competencia y cuáles no, para poder darles seguimiento con el sistema y monitorear su comportamiento. La información obtenida por este módulo es muy importante para la siguiente etapa del proyecto (Preparación del Dataset), ya que con ella puede prepararse un conjunto de datos perfectamente identificados de aquellos productos que son competencia y cuales no lo son.

5.3 Preparación del dataset

De acuerdo al análisis realizado se determinó que el usuario considera principalmente el precio del producto, la reputación del vendedor, el tipo de publicación, si ofrece envío gratis y si acepta pagos en mensualidades, para considerarlo como competencia y entonces monitorear dicho producto.

Una vez que se han identificado los atributos que serán utilizados durante el análisis, el siguiente paso es obtener y preparar los datos con los que se van a trabajar, para ello se toma una muestra de los registros que se tienen almacenados en la base datos, cabe mencionar que entre mayor sea el número de registros con los que se entrene el modelo los resultados serían más precisos. En este caso se ha tomado una muestra de 570 registros (Figura 71).

MELI ID	SELLER CUSTOM FIELD	TITLE	PRICE	FREE SHIPPING	MERCADOPAGO	LOGISTIC TYPE	LISTING TYPE ID
MLM784769354	DUMMY_272	Dummy Collector Limited Edition #272	7.0	✘	✔	drop_off	gold_special
MLM784769361	DUMMY_155	Dummy Collector Limited Edition #155	7.0	✘	✔	drop_off	gold_special
MLM784769388	DUMMY_216	Dummy Collector Limited Edition #216	7.0	✘	✔	drop_off	gold_special
MLM784769381	DUMMY_161	Dummy Collector Limited Edition #161	7.0	✘	✔	drop_off	gold_special
MLM784769373	DUMMY_226	Dummy Collector Limited Edition #226	7.0	✘	✔	drop_off	gold_special

Figura 71 Datos de registros almacenados en la base de datos sin transformar

El siguiente paso es preparar los datos para el análisis, para ello deben realizarse algunas transformaciones en los datos previamente con las cuales se podrán aplicar los algoritmos de clasificación correspondientes. La transformación consiste en convertir el valor que se obtuvo de Mercado Libre y que está almacenado en la base de datos, el cual generalmente se encuentra en formato tipo cadena de texto, para convertirlo a un formato numérico o booleano el cual puede ser utilizado por el algoritmo de clasificación, un ejemplo, es el precio del producto del competidor, de manera individual no nos ayuda de nada pero si se compara con el precio del producto del usuario se obtiene un nuevo valor que determina si el precio del producto del competidor es mejor que el del usuario.

Tomando la información anterior como referencia se asignan los valores 0, 1 y 2 al atributo del precio.

- El valor 0: si son iguales.
- El valor 1: si el precio del competidor es menor que el del usuario
- El valor 2: en caso que el precio del usuario es menor que el del competido.

La reputación del usuario se define en base a algunos criterios definidos por Mercado Libre como la atención del vendedor, la cantidad de ventas, cancelaciones y/o reclamos por parte de los compradores (Figura 72) la cual tiene un valor tipo cadena que hace referencia a un color que indica que tan confiable es un vendedor, en total son 5 colores, rojo, naranja, amarillo, verde claro y verde intenso, siendo el rojo el de menor reputación y verde el de mayor, para la reputación de la figura siguiente '3_yellow' indica que su reputación es intermedia.

```
"seller_reputation": {
  "level_id": "3_yellow",
  "power_seller_status": null,
  "transactions": {
    "canceled": 2142,
    "completed": 30884,
    "period": "historic",
    "ratings": {
      "negative": 0.03,
      "neutral": 0.02,
      "positive": 0.95
    },
    "total": 33026
  }
},
```

Figura 72 Datos de reputación del vendedor obtenidos de la plataforma en formato JSON

Para transformar estos se cambia el valor de la reputación y se les asignan los valores 1, 2, 3, 4 y 5 respectivamente. Posteriormente se hace compara el valor de la puntuación del usuario con la del vendedor y se asignan los siguientes valores:

- El valor 0: si las puntuaciones de ambos son iguales.
- El valor 1: si la puntuación del competidor es más alta que la del usuario.
- El valor 2: si la puntuación del competidor es más baja que la del usuario.

Para los atributos: tipo de publicación, envío gratis y pago en mensualidades, se asignan los valores 0 y 1, donde el 1 indica si se ofrece ese servicio y 0 en caso contrario. Por último, se define el atributo de la clase, el cual es lo que se busca clasificar, es decir, si es competencia o no, para este caso toma los valores 1 y 2, en caso de ser competencia toma el 1 y en caso contrario toma el 2. En el Cuadro 4 se puede observar un ejemplo de una parte de los datos transformados que serán utilizados en el proceso de análisis siguiente.

Calificación (R)	Mensualidades (P)	Precio (B)	Envío (E)	Publicación (T)	Competencia (C)
1	1	1	1	1	1
0	2	2	1	1	2
2	2	2	2	1	2
2	2	2	1	1	2
0	1	1	1	2	1
1	2	1	1	2	1
1	1	2	1	2	1
0	1	1	2	2	1
0	2	2	2	1	2
1	2	2	2	1	2
1	2	1	1	1	2
2	1	2	2	2	1
0	1	1	1	2	1
1	2	1	1	2	1
1	2	1	2	2	2
2	1	1	1	2	1
2	1	1	2	2	1

Tabla 4 Datos de entrenamiento para el modelo de clasificación

5.4 Definición del Modelo

En esta etapa se aplica el algoritmo de clasificación Naive Bayes para detectar los registros que son considerados competencia sobre los datos previamente obtenidos y procesados durante las etapas previas, para ello, se han utilizado las librerías de Python llamadas Scikit Learn y Pandas (Figura 73), las cuales sirven para

facilitar el análisis y procesamiento de datos, así como también permiten utilizar los diferentes algoritmos para minería de datos.



Figura 73 Librerías de Python Scikit Learn y Pandas

Para utilizar las librerías mencionadas, estas se deben agregar en la parte superior del archivo e indicar los módulos que se van a usar de cada librería durante el desarrollo (Figura 74).

```
import pandas as pd
from sklearn import datasets
from sklearn.model_selection import train_test_split
from sklearn.naive_bayes import GaussianNB
from sklearn.metrics import confusion_matrix, precision_score
```

Figura 74 Importación de librerías y módulos Scikit Learn y Pandas

Para trabajar con los datos obtenidos anteriormente se ha generado un archivo en formato CSV el cual se importa por medio de la instrucción `read_csv` de la librería pandas para poder procesarlo (Figura 75).

```
# Se carga el Dataset
data = pd.read_csv("competencia.csv")
print(data)
```

Figura 75 Definición del Dataset con Pandas

Una vez obtenido el archivo con los datos es momento de definir los atributos y la clase a utilizar durante el análisis, para ello se deben definir 2 dimensiones en los datos, una para los atributos y otra para la clase (Figura 76).

```
df = pd.DataFrame(data)

# Se definen los atributos y la Clase
x = df[['calificacion', 'pago', 'precio', 'envio', 'tipo']]
y = df['competencia']
```

Figura 76 Definición de Atributos y Clase en el Dataset

El siguiente paso es comenzar a entrenar el modelo, para ello se asigna la proporción de los datos que serán utilizados para entrenamiento y pruebas, para ello se utiliza un módulo de la librería Scikit Learn llamado *train_test_split* (Figura 77) en donde se indica que el 20% de los datos serán utilizados para pruebas y el 80% restante para el entrenamiento, las proporciones utilizadas pueden variar, no son valores específicos, pero debemos considerar tener suficientes datos para mejorar el desempeño del modelo.

```
# Se define la proporción para los datos de prueba y entrenamiento
X_train, X_test, y_train, y_test = train_test_split(x, y, test_size=0.2)

print("X_train")
print(X_train)
print("y_train")
print(y_train)
```

Figura 77 Asignación de Datos de Entrenamiento y Pruebas

Para continuar se define el algoritmo a utilizar, en este caso el Naive Bayes. Para ello se hace uso del método *GaussianNB* (Figura 78) importado previamente de la librería de Scikit Learn. Una vez definido el algoritmo, se entrena el modelo proporcionando los datos asignados para el entrenamiento.

```
# Se define el modelo a utilizar
algoritmo = GaussianNB()

# Se entrena el modelo
algoritmo.fit(X_train, y_train)
```

Figura 78 Entrenamiento del Modelo de Clasificación

5.4.1 Validación del Modelo

Una vez entrenado el modelo se realiza una predicción con la cual se evalúan los registros designados como prueba (el 20% del total de los registros utilizados) y se determina el desempeño del modelo. Para realizar la predicción se utiliza el método *predict* del algoritmo seleccionado (Figura 79).

```
# Una vez entrenado el modelo se hace una predicción
# utilizando los datos de prueba
y_pred = algoritmo.predict(X_test)
```

Figura 79 Predicción de datos de prueba

Posteriormente a haber realizado la predicción con los datos de prueba se deben de obtener las métricas de desempeño del modelo generando una matriz de confusión y calcular el score de precisión, en caso de que los valores de las métricas obtenidas sean demasiado bajas, se debe repetir el entrenamiento hasta alcanzar resultados favorables (Figura 80).

```
# Se obtiene la matriz de confusión para determinar
# la efectividad del algoritmo
matriz = confusion_matrix(y_test, y_pred)
print("matriz de confusión")
print(matriz)

# Se obtiene la precisión del algoritmo
precision = precision_score(y_test, y_pred)
print(precision)
```

Figura 80 Obtención de métricas de desempeño del modelo

La librería de *Scikit Learn* cuenta con un módulo para obtener las métricas para determinar el desempeño del modelo, llamado *sklearn.metrics* el cual fue importado anteriormente (Figura 35). Este módulo dispone de un conjunto de funciones para calcular estos valores, para obtener la matriz de confusión se utiliza el método llamado *confusion_matrix* el cual recibe como parámetros de entrada los datos que fueron utilizados para las pruebas y los resultados de la predicción, el resultado de esta función es una matriz cuadrada que indica la cantidad de aciertos logrados, así como también la cantidad de errores obtenidos (Figura 81).

	Competencia	No Competencia	
Competencia	37	3	40
No Competencia	2	72	74
	39	75	

Figura 81 Matriz de confusión para la clase Competencia

El score de precisión puede obtenerse llamando al método *precision_score* de la misma librería, el cual recibe como parámetros de entrada los datos de prueba y el resultado de la predicción para calcular el porcentaje de eficiencia del modelo (Figura 82).

Nombre	Tipo	Tamaño	Valor
X	float64	(569, 30)	[[1.799e+01 1.038e+01 1.228e+02 ... 2.654e-01 4.601e-01 1.189e-01] [2 ...
X_test	float64	(114, 30)	[[1.265e+01 1.817e+01 8.269e+01 ... 1.407e-01 3.230e-01 1.033e-01] [1 ...
X_train	float64	(455, 30)	[[1.278e+01 1.649e+01 8.137e+01 ... 5.882e-02 2.383e-01 6.410e-02] [1 ...
Y	int32	(569,)	[0 0 0 ... 0 0 1]
data	utils.Bunch	6	Bunch object of sklearn.utils module
df	DataFrame	(14, 6)	Column names: calificacion, pago, precio, envio, tipo, competencia
matriz	int64	(2, 2)	[[39 3] [3 69]]
precision	float64	1	0.9583333333333334
x	DataFrame	(14, 5)	Column names: calificacion, pago, precio, envio, tipo
y	Series	(14,)	Series object of pandas.core.series module
y_pred	int32	(114,)	[1 1 1 ... 0 0 0]
y_test	int32	(114,)	[1 1 1 ... 0 0 0]
y_train	int32	(455,)	[1 1 0 ... 0 1 1]

Figura 82 Score de precisión del modelo de clasificación

5.5 Detección automática

En esta etapa se analizan los resultados obtenidos por el algoritmo aplicado en la etapa anterior, para ello se propone el desarrollo de un módulo dentro del sistema que permita visualizar mediante un reporte los datos de los productos que hayan sido detectados por el algoritmo. Como puede verse a continuación, el reporte debe mostrar el precio de los productos e indicar si son mayores o menores que el precio del producto del usuario (Figura 83).

El usuario podrá determinar si el resultado de la detección ha sido correcto, y en caso contrario indicar en el sistema si el producto no se considera competencia, de modo que el modelo de clasificación vaya mejorando conforme se utilice.

Producto	Precio	
 Ibanezguitarra Electrica Rga Azul Sombreada Rga42fm EQUIPO PROFESIONAL MLM650244622 Categoría: MLM-ELECTRIC_GUITARS Envío: DROP_OFF Tipo: GOLD_PRO	\$8,290.00 Anterior: \$8,290.00	Activa Update Remove 2019-06-13 13:24
 Guitarra Ibanez Rga42fm-blf Nueva SOLODEBATERIA MLM638029720 Categoría: MLM-ELECTRIC_GUITARS Vendidos: 1 de 1 Envío: DROP_OFF Tipo: GOLD_PRO	\$8,440.00 Anterior: \$8,440.00	Pausada

MLM590290545
 Guitarra Eléctrica Ibanez Rga, Azul Somb. Rga42fm Blf Tipo: PRO Envío: DROP_OFF Precio actual: \$8,299.00 Precio Sugerido: \$8,290.00 Antes: \$8,299.00

Figura 83 Reporte de Resultados, Seguimiento de Competencia y Precio Sugerido

En esta etapa se plantea que el sistema realice el seguimiento automático de los productos que se hayan marcado como competencia por el algoritmo de clasificación.

En esta parte, el sistema debe de hacer consultas cada cierto periodo de tiempo al API de Mercado Libre con el fin de obtener los datos actualizados de los productos y actualizar la información registrada en la base de datos del sistema, en caso de detectar cambios en las publicaciones de otros vendedores, el sistema generará una alerta para el usuario indicando por ejemplo un precio sugerido para su producto en caso que los productos de la competencia sean mejores, lo anteriormente mencionado se puede observar en la Figura 9 donde se puede ver un producto con diferentes precios, entre ellos el actual y el sugerido.

5.6 Preparación del ambiente de producción

El siguiente paso en el proyecto, una vez concluido el desarrollo, es pasar el proyecto a un ambiente productivo y montarlo en un servidor, para ello se deben realizar algunas configuraciones previas y preparar el servidor en donde estará alojado. Estas configuraciones se detallan en los siguientes apartados.

5.6.1 Creación del servidor virtual privado (VPS)

El servidor en donde se va a montar el ambiente productivo será un VPS el cual permite rentar un espacio dedicado en la nube para poder montar aplicaciones. El VPS puede ser configurado con el sistema operativo se desee y con los recursos que se requieran en cuanto a memoria y procesamiento se refieren, no es necesario realizar inversiones económicas en un inicio ya que es un servicio bajo demanda y se paga solo por lo que se utilice de forma mensual. El proveedor del servicio VPS utilizado es Digital Ocean ya que ofrece el mejor servicio en relación a cantidad de recursos y precio, además de que es muy fácil de utilizar y cuenta con una gran aceptación por parte de la comunidad de desarrolladores (Figura 84).

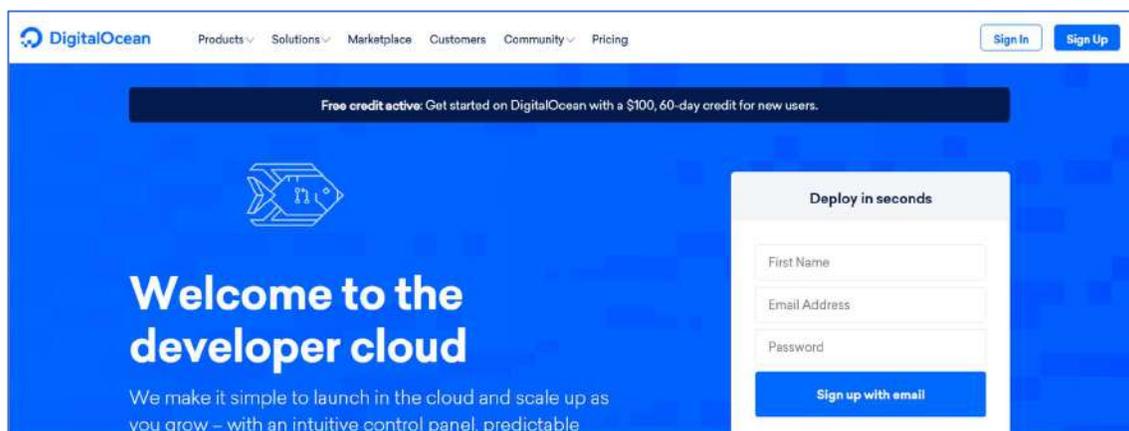


Figura 84 Proveedor de servicios VPS Digital Ocean

Digital Ocean con diferentes planes de servidores que permiten elegir el que mejor se adapte a la aplicación desarrollada y da la oportunidad de escalar los recursos al momento que estos se requieran (Figura 85).

Basic Droplets ⓘ

Balanced virtual machines with a healthy amount of memory tuned to host and scale applications like blogs, web apps, testing and staging environments, in-memory caching, and databases. [Learn more](#)

Memory	vCPUs	Transfer	SSD Disk	\$/HR	\$/MO	
1GB	1vCPU	1TB	25GB	\$0.007	\$5	Sign up
2GB	1vCPU	2TB	50GB	\$0.015	\$10	Sign up
2GB	2vCPUs	3TB	60GB	\$0.022	\$15	Sign up
4GB	2vCPUs	4TB	80GB	\$0.030	\$20	Sign up
8GB	4vCPUs	5TB	160GB	\$0.060	\$40	Sign up
16GB	8vCPUs	6TB	320GB	\$0.119	\$80	Sign up

Figura 85 Planes de VPS ofrecidos por Digital Ocean

Es posible elegir entre diferentes sistemas operativos para pasar el proyecto a producción, ya que el servicio ofrece imágenes para todas las distribuciones de Linux como son Ubuntu, FreeBSD, Fedora, Debian y CentOS, permitiendo elegir en cada una de ellas la versión del SO que se desee (Figura 86).

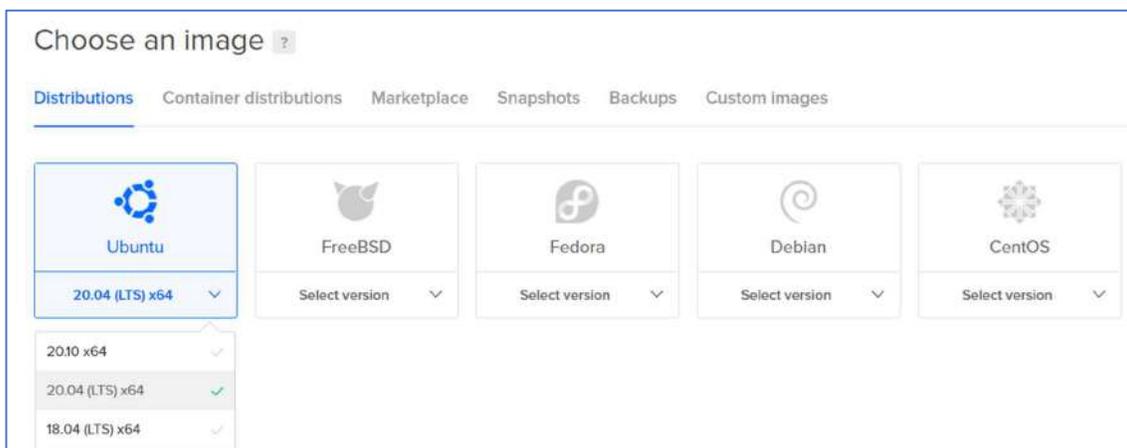


Figura 86 Diferentes imágenes de sistemas operativos ofrecidos por Digital Ocean

Para este proyecto se ha decidido comenzar con un servidor configurado con Linux Ubuntu versión 18.0 (LTS) x64 utilizando 2 CPU y 4 GB de RAM por un precio de 15 USD mensuales. En un principio se había elegido un plan más económico, sin embargo, el tiempo de respuesta no era suficiente, por lo que se decidió incrementarlo (Figura 87).

Choose a plan Help me choose

SHARED CPU | DEDICATED CPU

Basic | General Purpose | CPU-Optimized | Memory-Optimized | Storage-Optimized **NEW**

Basic virtual machines with a mix of memory and compute resources. Best for small projects that can handle variable levels of CPU performance, like blogs, web apps and dev/test environments.

Price	Resources
\$5/mo \$0.007/hour	1 GB / 1 CPU 25 GB SSD Disk 1000 GB transfer
\$10/mo \$0.015/hour	2 GB / 1 CPU 50 GB SSD Disk 2 TB transfer
\$15/mo \$0.022/hour	2 GB / 2 CPUs 60 GB SSD Disk 3 TB transfer
\$20/mo \$0.030/hour	4 GB / 2 CPUs 80 GB SSD Disk 4 TB transfer
\$40/mo \$0.060/hour	8 GB / 4 CPUs 160 GB SSD Disk 5 TB transfer
\$80/mo \$0.119/hour	16 GB / 8 CPUs 320 GB SSD Disk 6 TB transfer

Figura 87 Plan seleccionado para poner el proyecto en producción

Una vez creado el servidor (Droplet), Digital Ocean le asigna una IP publica al servidor del proyecto, así como también un usuario y password mediante el cual se puede realizar una conexión SSH¹ y acceder al servidor desde una terminal (Figura 88).

promusicgear in Gerardo Alberto Lopez Ve... / 4 GB Memory / 50 GB Disk / SFO2 - Ubuntu odoo11-error.js ON

ipv4: 167.99160.33 | ipv6: Enable now | Private IP: No VPC Network | Floating IP: Enable now | Console:

Graphs | Access | Power | Volumes | Resize | Networking | Backups | Snapshots | Kernel | History | Destroy | Tags | Recovery

NEW! Upgrade your Droplet for additional metrics and alerting. [Learn How to Update](#)

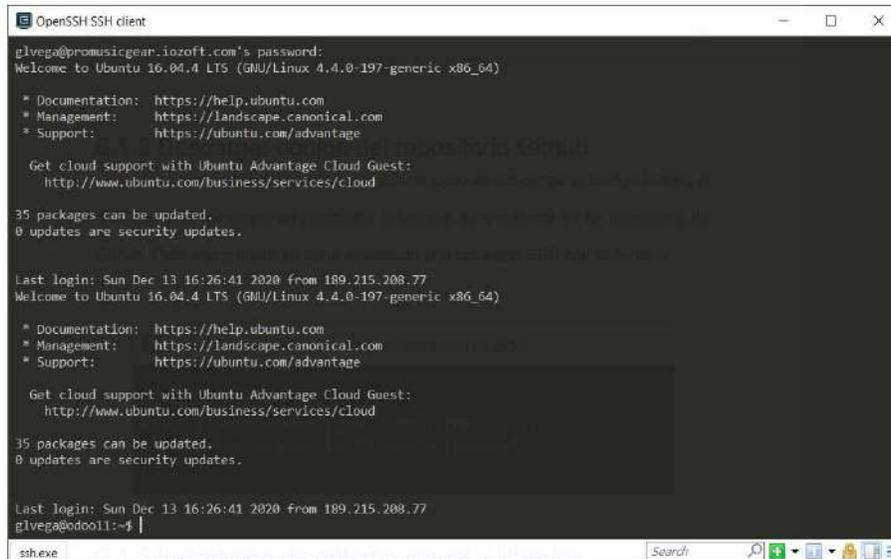
Bandwidth (Last 1 hour)

Time	Bandwidth (kb/s)
11:10	~30
11:15	~10
11:20	~10
11:25	~10
11:30	~10
11:35	~10
11:40	~70
11:45	~70
11:50	~20
11:55	~20
12:00	~20
12:05	~10

¹ SSH es el nombre de un protocolo y del programa que lo implementa cuya principal función es el acceso remoto a un servidor por medio de un canal seguro en el que toda la información está cifrada.

5.6.2 Descargar código del repositorio Github

Una vez creado el servidor el siguiente paso es descargar el código fuente, el cual como se mencionó en apartados anteriores se encuentra en un repositorio de Github. Para ello primero se debe establecer una conexión SSH con el servidor (Figura 89).



```
OpenSSH SSH client
glvega@promusicgear.iozoft.com's password:
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.4.0-197-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

35 packages can be updated.
0 updates are security updates.

Last login: Sun Dec 13 16:26:41 2020 from 189.215.208.77
Welcome to Ubuntu 16.04.4 LTS (GNU/Linux 4.4.0-197-generic x86_64)

 * Documentation:  https://help.ubuntu.com
 * Management:    https://landscape.canonical.com
 * Support:       https://ubuntu.com/advantage

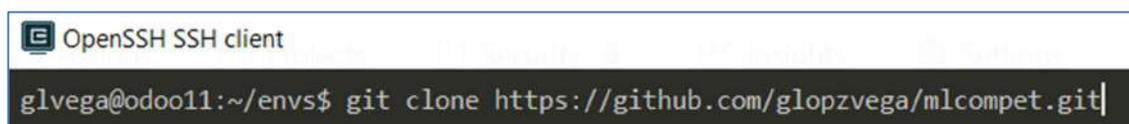
Get cloud support with Ubuntu Advantage Cloud Guest:
http://www.ubuntu.com/business/services/cloud

35 packages can be updated.
0 updates are security updates.

Last login: Sun Dec 13 16:26:41 2020 from 189.215.208.77
glvega@odoo11:~$
```

Figura 89 Conexión SSH al servidor del proyecto

Una vez dentro del servidor, seleccionamos una ubicación en donde se colocara el código fuente descargado, para el caso de este proyecto se ha elegido la carpeta /home en donde se ha creado una carpeta con el nombre /envs en donde se colocara el proyecto. Una vez creada la carpeta ejecutamos el comando git clone para clonar y descargar el repositorio de Github, para ello se debe indicar en el comando la URL del repositorio (Figura 90).



```
OpenSSH SSH client
glvega@odoo11:~/envs$ git clone https://github.com/glopzvega/mlcompet.git
```

Figura 90 Clonar repositorio del proyecto en Github

La URL del repositorio puede obtenerse desde el sitio de Github (Figura 91).

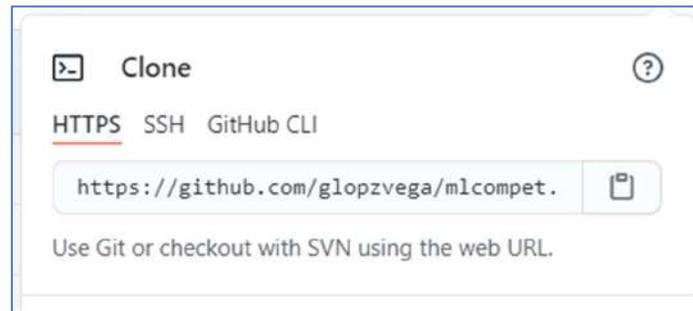


Figura 91 URL repositorio del proyecto en Github

Una vez clonado el repositorio se habrá descargado el código fuente del proyecto, el cual podremos examinar en la consola ejecutando el siguiente comando (Figura 92).

```
OpenSSH SSH client
glvega@odoo11:~/envs/pmgenv/mlcompet$ ls
accounts  comentarios  dataset.csv  manage.py  mlcompet  odoo  Procfile  requirements.txt  static
buscador  datadump.json  db.sqlite3  meli_conn  mlcompet.log  preguntas  README.md  respuestas  templates
glvega@odoo11:~/envs/pmgenv/mlcompet$ ls -l
total 481488
drwxr-xr-x 7 root root 4096 Dec 5 18:34 accounts
drwxr-xr-x 6 root root 4096 Dec 5 18:34 buscador
drwxr-xr-x 6 root root 4096 Nov 30 04:09 comentarios
-rw-r--r-- 1 root root 80332119 Nov 30 04:09 datadump.json
-rw-r--r-- 1 root root 7920 Nov 30 04:09 dataset.csv
-rw-r--r-- 1 root root 411619328 Dec 5 18:26 db.sqlite3
-rw-r--r-- 1 root root 540 Oct 28 2018 manage.py
drwxr-xr-x 6 root root 4096 Nov 30 04:09 meli_conn
drwxr-xr-x 3 root root 4096 Dec 5 18:31 mlcompet
-rw-r--r-- 1 root root 1011481 Dec 13 18:06 mlcompet.log
drwxr-xr-x 4 root root 4096 Nov 30 04:09 odoo
drwxr-xr-x 6 root root 4096 Nov 30 04:09 preguntas
-rw-r--r-- 1 root root 41 Oct 28 2018 Procfile
-rw-r--r-- 1 root root 10 Oct 28 2018 README.md
-rw-r--r-- 1 root root 264 Nov 30 04:09 requirements.txt
drwxr-xr-x 6 root root 4096 Nov 30 04:09 respuestas
drwxr-xr-x 17 root root 4096 Dec 5 19:01 static
drwxr-xr-x 3 root root 4096 Nov 30 04:10 templates
glvega@odoo11:~/envs/pmgenv/mlcompet$
```

Figura 92 Código fuente descargado en el servidor

5.6.3 Instalación de entorno virtual y dependencias

Una vez descargado el código fuente, el siguiente paso es instalarlo, para ello es necesario preparar el entorno virtual e instalar las dependencias del proyecto. Las dependencias del proyecto están guardadas en el archivo `requirements.txt` el cual se generó en apartados anteriores.

Para instalar las dependencias se debe crear en primera instancia el entorno virtual con ayuda del comando `python -m venv entornovirtual`, una vez creado procedemos a activarlo con el comando `source /bin/activate`. Una vez hecho esto, se debe haber creado una carpeta con el nombre que se indicó en el comando, la cual deberá tener una estructura similar a la que se muestra a continuación (Figura 93).

```
glvega@odoo11:~/envs/pmgenv$ ls -l
total 28
drwxr-xr-x  3 root root 4096 Oct 28  2018 bin
drwxr-xr-x  2 root root 4096 Oct 28  2018 include
drwxr-xr-x  3 root root 4096 Oct 28  2018 lib
lrwxrwxrwx  1 root root    3 Oct 28  2018 lib64 -> lib
drwxr-xr-x 14 root root 4096 Dec  5 18:26 mlcompet
srwxrwxrwx  1 root root    0 Dec 13 16:27 mlcompet.sock
-rw-r--r--  1 root root  59 Oct 28  2018 pip-selfcheck.json
-rw-r--r--  1 root root  69 Oct 28  2018 pyvenv.cfg
drwxr-xr-x  3 root root 4096 Oct 28  2018 share
glvega@odoo11:~/envs/pmgenv$ |
```

Figura 93 Entorno virtual creado en servidor de producción

Una vez creado y activado el entorno virtual, el siguiente paso es instalar las dependencias, para ello se debe acceder a la carpeta del proyecto y ejecutar el comando `pip install -r requirements.txt` el cual instalara una a una las dependencias del proyecto (Figura 94).

```
OpenSSH SSH client
(pmgenv) glvega@odoo11:~/envs/pmgenv/mlcompet$ cd ..
(pmgenv) glvega@odoo11:~/envs/pmgenv$ cd ..
(pmgenv) glvega@odoo11:~/envs$ cd pmgenv/mlcompet/
(pmgenv) glvega@odoo11:~/envs/pmgenv/mlcompet$ pip install -r requirements.txt
```

Figura 94 Instalación de dependencias del proyecto

Una vez ejecutado el comando anterior, se comenzarán a descargar e instalar las librerías que se encuentran dentro del archivo de requirements.txt, una vez finalizada la instalación ya se tendrán listas todas las dependencias para comenzar a utilizar el proyecto (Figura 95).

```
OpenSSH SSH client
glvega@promusicgear.iozsoft.com's password:
(pmgenv) glvega@odoo11:~/envs/pmgenv/mlcompet$ cd ..
(pmgenv) glvega@odoo11:~/envs/pmgenv$ cd ..
(pmgenv) glvega@odoo11:~/envs$ cd pmgenv/mlcompet/
(pmgenv) glvega@odoo11:~/envs/pmgenv/mlcompet$ pip install -r requirements.txt
Collecting asgiref==3.2.10 (from -r requirements.txt (line 1))
  Downloading https://files.pythonhosted.org/packages/d5/eb/64725b25f991010307fd18a9e0c1f0e6dff2f03622fc4bcbcd2244f60d6/asgiref-3.2.10-py3-none-any.whl
Requirement already satisfied (use --upgrade to upgrade): certifi==2018.8.24 in /home/glvega/envs/pmgenv/lib/python3.5/site-packages (from -r requirements.txt (line 2))
Requirement already satisfied (use --upgrade to upgrade): chardet==3.0.4 in /home/glvega/envs/pmgenv/lib/python3.5/site-packages (from -r requirements.txt (line 3))
Requirement already satisfied (use --upgrade to upgrade): dj-database-url==0.5.0 in /home/glvega/envs/pmgenv/lib/python3.5/site-packages (from -r requirements.txt (line 4))
Collecting Django==3.1.2 (from -r requirements.txt (line 5))
  Downloading https://files.pythonhosted.org/packages/50/22/4c91847becadbb54b5a518909ed5000bb1777168c7d6b087e8f79e5e05b/Django-3.1.2-py3-none-any.whl (7.8MB)
100% |#####| 7.8MB 135kB/s
Requirement already satisfied (use --upgrade to upgrade): django-common-helpers==0.9.2 in /home/glvega/envs/pmgenv/lib/python3.5/site-packages (from -r requirements.txt (line 6))
Requirement already satisfied (use --upgrade to upgrade): django-cron==0.5.1 in /home/glvega/envs/pmgenv/lib/python3.5/site-packages (from -r requirements.txt (line 7))
Collecting gunicorn==20.0.4 (from -r requirements.txt (line 8))
  Downloading https://files.pythonhosted.org/packages/69/ca/926f7cd3a2014b16870086b2d0fdc84a9e49473c68a8dff8b57f7c156f43/gunicorn-20.0.4-py2.py3-none-any.whl (77kB)
100% |#####| 81kB 9.6MB/s
Collecting idna==2.10 (from -r requirements.txt (line 9))
  Downloading https://files.pythonhosted.org/packages/a2/38/928ddce2273eaa564f6f50de919327bf3a00f091b5baba8dfa9460f3a8a8/idna-2.10-py2.py3-none-any.whl (58kB)
100% |#####| 61kB 6.2MB/s
ssh.exe Search
```

Figura 95 Instalación de dependencias del proyecto.

5.6.4 Configuración de Nginx y Gunicorn

Cuando se pasa de un ambiente de desarrollo a producción ya no puede utilizarse el comando de Django `python manage.py runserver`, el cual es un servidor web para ambientes de desarrollo, pero no es recomendable para servir el proyecto a través de un servidor web en ambientes productivos, esto debido a que no es robusto y tiene algunos problemas de seguridad, en su lugar deben configurarse la librería Gunicorn, la cual es una interfaz de puerta de enlace de servidor web en Python. Para configurarla se deben realizar algunas modificaciones en algunos archivos del servidor.

La librería Gunicorn es una de las dependencias que se incluyeron en el archivo de requerimientos, por lo que ya se debe encontrar instalada en el servidor de producción después de ejecutar los comandos del apartado anterior. Una vez instalada, se puede servir el proyecto ejecutando el siguiente comando (Figura 96).

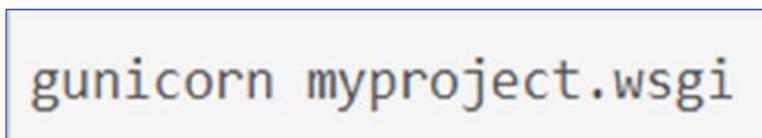
A terminal window with a light gray background and a blue border. The text 'gunicorn myproject.wsgi' is displayed in a monospaced font, with 'gunicorn' in blue and 'myproject.wsgi' in black.

Figura 96 Iniciar servidor web del proyecto con Gunicorn

El archivo WSGI² es un archivo que se encuentra dentro del proyecto, el cual se genera al momento de crear cualquier proyecto en Django Framework, se encuentra en la carpeta que tiene el mismo nombre del proyecto, junto al archivo de configuración (Figura 97).

² Web Server Gateway Interface

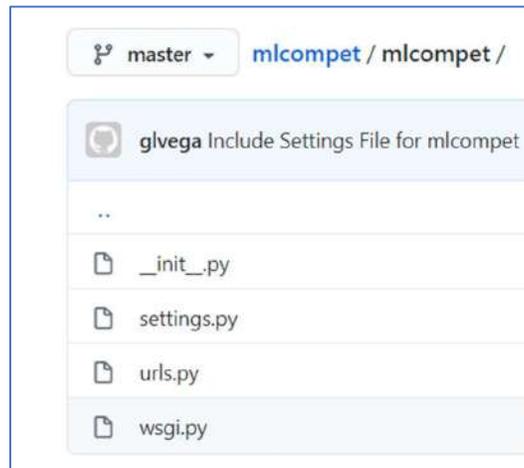


Figura 97 Archivo WSGI del proyecto

Para servir proyectos web en producción debe utilizarse este archivo, ya que de este modo permite servir un proyecto de una forma robusta, permitiendo así múltiples conexiones, resolviendo a su vez diferentes problemas de seguridad.

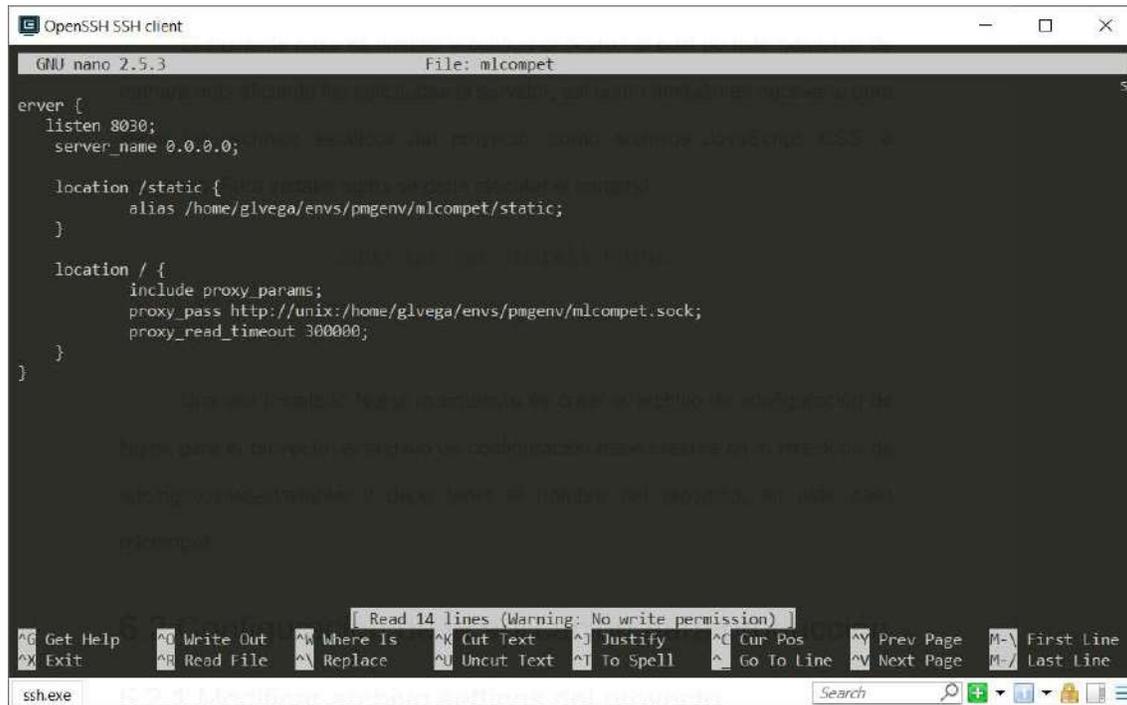
El siguiente paso es instalar y configurar Nginx³ el cual permite gestionar de manera más eficiente las solicitudes al servidor, así como también es necesario para servir los archivos estáticos del proyecto, como archivos JavaScript, CSS, e imágenes. Para instalar nginx se debe ejecutar el comando siguiente (Figura 98).

```
sudo apt-get install nginx
```

Figura 98 Instalación de Nginx

Una vez instalado Nginx lo siguiente es crear el archivo de configuración de Nginx para el proyecto, el archivo de configuración debe crearse en el directorio de /etc/nginx/sites-available/ y debe tener el nombre del proyecto, en este caso mlcompet (Figura 99).

³ Nginx es un servidor web/proxy inverso ligero de alto rendimiento y un proxy para protocolos de correo electrónico.



```
GNU nano 2.5.3 File: mlcompet
server {
  listen 8030;
  server_name 0.0.0.0;

  location /static {
    alias /home/glvega/envs/pmgenv/mlcompet/static;
  }

  location / {
    include proxy_params;
    proxy_pass http://unix:/home/glvega/envs/pmgenv/mlcompet.sock;
    proxy_read_timeout 300000;
  }
}
```

Figura 99 Archivo de configuración de Nginx para el proyecto.

En el archivo se configura el puerto, IP, ubicación de los archivos estáticos, así como también la ubicación del archivo de Unicorn, entre otras.

5.6.5 Modificar archivo settings del proyecto

Una vez configurado el servidor web del proyecto, el siguiente paso es configurar el proyecto como tal para ser accedido desde ambientes productivos por medio de una IP pública o dominio, en lugar que desde localhost. Para ello se deben realizar algunos ajustes en el archivo de configuración del proyecto, por ejemplo en la sección de host permitidos, se debe incluir el dominio desde donde se puede acceder a la aplicación (Figura 30).



```
ALLOWED_HOSTS = [
    'localhost',
    'mlcompet.iozsoft.com'
]
```

Figura 100 Configuración de host permitidos al proyecto en producción

Otra configuración que debe realizarse es actualizar algunas de las variables que se utilizan dentro del proyecto, para hacer referencia a la nueva ubicación del proyecto (Figura 101).

```
SITE_URL = 'https://mlcompet.iozoft.com/'
# SITE_URL = 'http://localhost/'
```

Figura 101 Configuración de constantes utilizadas en el código

En el archivo de configuración, también se define la conexión a la base de datos que utilizara el proyecto, por default los proyectos desarrollados en Django Framework utilizan la base de datos sqlite3, la cual es una base de datos ligera que en combinación con Django permiten una adecuada gestión de los datos. Para este proyecto se ha utilizado esta base de datos, sin embargo, se planea cambiar a una base de datos más robusta para una siguiente versión del proyecto (Figura 102).

```
TEMPLATES = [
    {
        'BACKEND': 'django.template.backends.django.DjangoTemplates',
        'DIRS': [
            os.path.join(BASE_DIR, 'templates')
        ],
        'APP_DIRS': True,
        'OPTIONS': {
            'context_processors': [
                'django.template.context_processors.debug',
                'django.template.context_processors.request',
                'django.contrib.auth.context_processors.auth',
                'django.contrib.messages.context_processors.messages',
            ],
        },
    },
]

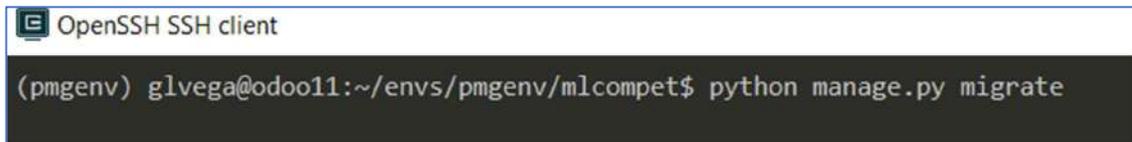
WSGI_APPLICATION = 'mlcompet.wsgi.application'

# Database
# https://docs.djangoproject.com/en/2.1/ref/settings/#databases

DATABASES = {
    'default': {
        'ENGINE': 'django.db.backends.sqlite3',
        'NAME': os.path.join(BASE_DIR, 'db.sqlite3'),
    }
}
```

5.6.6 Inicializar aplicación y base de datos

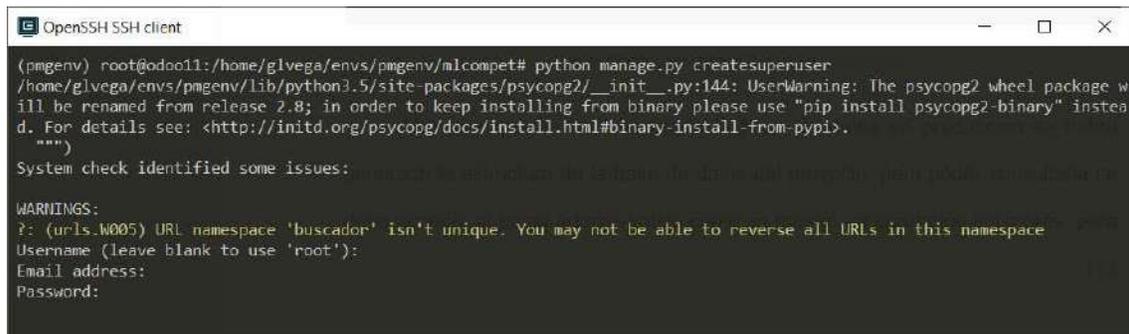
Después de la configuración del proyecto, el siguiente paso es inicializar el proyecto y la base de datos, para ello debemos tener configurada previamente la base de datos en archivo de configuración como se mostró en apartados anteriores. Una vez que se ha establecido la conexión con la base de datos se debe ejecutar el siguiente comando para inicializar la base de datos (Figura 103).



```
OpenSSH SSH client
(pmgenv) glvega@odoo11:~/envs/pmgenv/mlcompet$ python manage.py migrate
```

Figura 103 Inicializar base de datos en producción

Una vez que se ha inicializado la base de datos en producción se habrá generado la estructura de la base de datos del proyecto, para poder consultarla se debe acceder al panel administrativo como se mostró en apartados anteriores, para ello hay que crear un super usuario para acceder. Para crear el super usuario se ejecuta el siguiente comando, el cual solicita el nombre del usuario, password y email (Figura 104).



```
OpenSSH SSH client
(pmgenv) root@odoo11:/home/glvega/envs/pmgenv/mlcompet# python manage.py createsuperuser
/home/glvega/envs/pmgenv/lib/python3.5/site-packages/psycopy2/__init__.py:144: UserWarning: The psycopy2 wheel package will be renamed from release 2.8; in order to keep installing from binary please use "pip install psycopy2-binary" instead. For details see: <http://initd.org/psycopy/docs/install.html#binary-install-from-pypi>.
  """
System check identified some issues:

WARNINGS:
?: (urls.W005) URL namespace 'buscador' isn't unique. You may not be able to reverse all URLs in this namespace
Username (leave blank to use 'root'):
Email address:
Password:
```

Figura 104 Creación del super usuario en producción

Una vez creado el super usuario se podrá acceder panel de administración indicando la URL del proyecto agregando /admin al final, en donde solicitará los datos de accesos creados anteriormente (Figura 105).

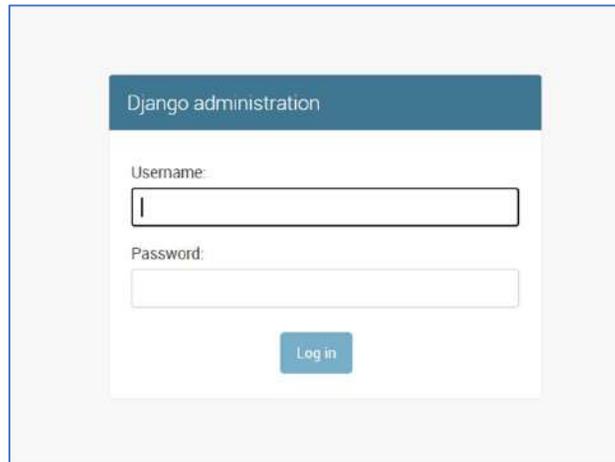
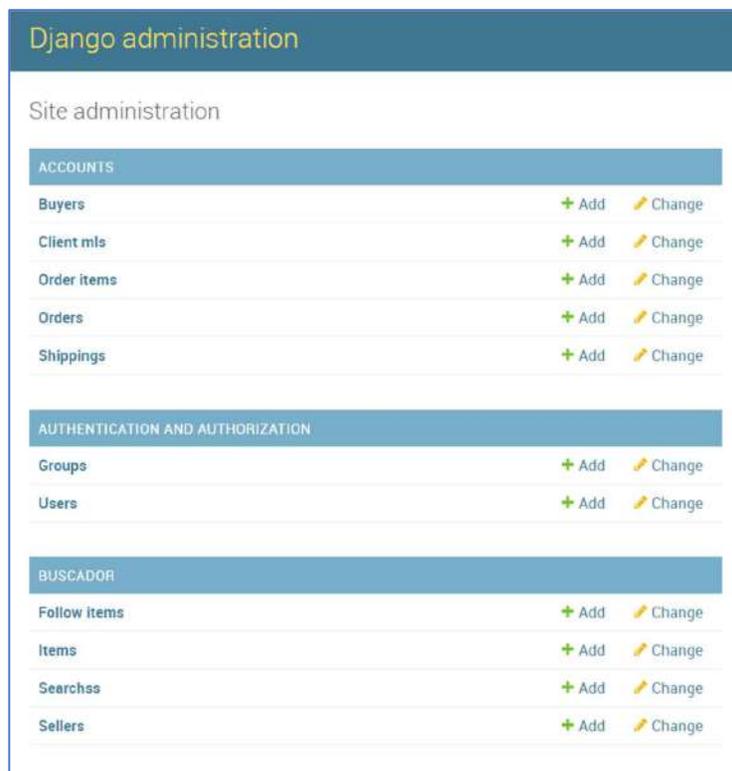


Figura 105 Acceso al panel administrativo en producción

Una vez dentro del panel administrativo ya se podrá consultar la estructura de la base de datos creada (Figura 106).



ACCOUNTS	
Buyers	+ Add Change
Client mls	+ Add Change
Order items	+ Add Change
Orders	+ Add Change
Shippings	+ Add Change
AUTHENTICATION AND AUTHORIZATION	
Groups	+ Add Change
Users	+ Add Change
BUSCADOR	
Follow items	+ Add Change
Items	+ Add Change
Searchss	+ Add Change
Sellers	+ Add Change

Figura 106 Base de datos inicializada en producción.

5.6.7 Configurar cuenta del usuario

Una vez inicializada la aplicación y la base de datos, el siguiente paso es crear la cuenta del usuario, para ello hay que acceder al panel administrativo del proyecto y seleccionar el modelo de usuarios del apartado de Autenticación y Autorización (Figura 107).



Figura 107 Apartado de autenticación y autorización del proyecto

Una vez dentro del apartado de usuarios, se selecciona la opción de agregar nuevo usuario, en donde se solicitará los datos de acceso, en este caso un usuario y password (Figura 108).

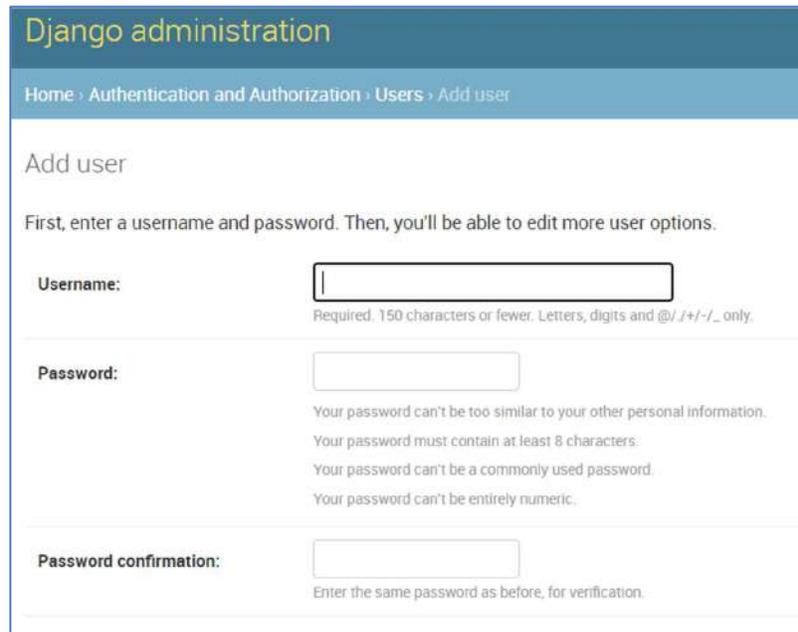
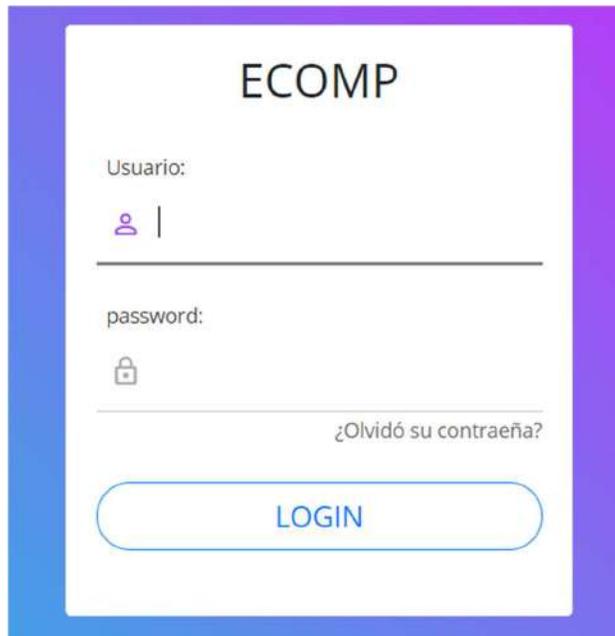
The image shows a screenshot of the Django administration 'Add user' form. The page title is 'Django administration' and the breadcrumb trail is 'Home > Authentication and Authorization > Users > Add user'. The form is titled 'Add user' and includes the instruction: 'First, enter a username and password. Then, you'll be able to edit more user options.' There are three input fields: 'Username:', 'Password:', and 'Password confirmation:'. The 'Username' field has a validation message: 'Required. 150 characters or fewer. Letters, digits and @/./+/-/_ only.' The 'Password' field has four validation messages: 'Your password can't be too similar to your other personal information.', 'Your password must contain at least 8 characters.', 'Your password can't be a commonly used password.', and 'Your password can't be entirely numeric.' The 'Password confirmation' field has a validation message: 'Enter the same password as before, for verification.'

Figura 108 Creación de cuenta de usuario.

Una vez creado el usuario, este aparecerá en el listado de usuarios y podrá acceder a la aplicación desde el formulario de acceso (Figura 109).



5.6.8 Conectar aplicación a Mercado Libre

Una vez dentro de la aplicación, lo primero que hay que hacer es vincular la aplicación con la cuenta de Mercado Libre, para ello se debe acceder a la sección de Perfil dentro del menú de la aplicación (Figura 109).



Figura 109 Menú Perfil de la aplicación

Una vez dentro del Perfil debe indicar que no se tiene vinculada ninguna cuenta de Mercado Libre, por lo que es necesario darle clic al enlace, el cual te enviara a la aplicación de la tienda en donde debes autorizar que la aplicación pueda acceder a los datos de tu cuenta (Figura 110).

Accede a MercadoLibre



El primer paso para comenzar a trabajar con mercadolibre es dar permisos a la aplicación para gestionar tus datos y publicaciones. Para ello vamos a dar click en el siguiente enlace.

[Conectarse a MercadoLibre](#)

The diagram features a yellow background with various icons: a smartphone, a shopping cart, a laptop, a gear, a mail icon, a checkmark, a cloud, and a pie chart. A central text reads 'integración con mercado libre'. The Mercado Libre logo is in the bottom left corner, and the word 'e-commerce' is written in the top right.

Figura 110 Conexión a Mercado Libre

Una vez hecho esto, te debe redirigir nuevamente a la aplicación en donde ya se podrá comenzar a hacer consultas (Figura 111).

Autorizar conexión



Hola Gerardo Alberto, autoriza a que la aplicación **"MLCOMPET2"** se conecte con Mercado Libre

Podrá ver datos privados de tu cuenta. ⓘ

Podrás gestionar tus autorizaciones desde la Herramienta de Seguridad en tu cuenta.

The authorization screen shows a grey box icon on the left and a handshake icon on the right, connected by two arrows. The text is centered and includes a question mark icon for help.

Figura 111 Autorización de acceso a la aplicación a la cuenta de Mercado Libre

Una vez autorizado el acceso debe mostrarse en la aplicación un token de conexión el cual es utilizado para realizar las consultas a Mercado Libre y que significa que la cuenta se ha vinculado con éxito (Figura 112).



Figura 112 Cuenta vinculada a Mercado Libre

5.6.9 Obtener catálogo de productos del usuario

Una vez vinculada la cuenta, ya se pueden realizar búsquedas a través de la aplicación, sin embargo, para poder realizar el análisis y seguimiento de productos de la competencia, es necesario descargar el catálogo de productos del usuario, ya que estos servirán de referencia para determinar si los productos de otros vendedores son más competitivos que los del usuario.

Para descargar el catálogo del usuario debe ir al apartado de Mis productos, en donde el usuario podrá filtrar los productos que desea descargar introduciendo el nombre o palabras claves con lo que la búsqueda se hará en base a ellas, o en su defecto descargar todo el catálogo (Figura 113).



Figura 113 Búsqueda de productos en el catálogo del usuario

Una vez hecha la búsqueda, se listará en la parte inferior de la ventana los productos obtenidos desde Mercado Libre, los cuales, para este proyecto son 11,998 registros (Figura 114).

Total : 50 de 11998, Página: 1 de 240

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24 25 26 27

Imagen	Producto	Precio	Estatus	Odoos
	100 Plumillas El Cometa Para Guitarra Celuloide 110 ✓ MLM685629334 ✓ NA ✓ FREE Mercado	\$1039.00	✓ ESPECIAL	Editar ✓ 65182
	100 Plumillas El Cometa Para Guitarra Celuloide 110 ✓ MLM685629447 ✓ NA ✓ FREE Mercado	\$1119.00	✓ PRO	Editar ✓ 65182
	10 Aceites Jupiter Jcm-ko2 Para Llaves De Instrumentos De Viento ✓ MLM709220028 ✓ PMG9346 ✓ FREE Mercado	\$1953.00	PAUSADO ✓ ESPECIAL	Editar ✓ 64652

Figura 114 Catálogo de productos del usuario

Una vez realizada la importación de productos del catalogo del usuario, queda concluida la etapa de implementación del proyecto y el usuario ya puede utilizar el sistema para realizar las consultas y el seguimiento de los productos de su competencia, en los apartados siguientes se analizan los resultados obtenidos del presente proyecto, en donde se observan los módulos desarrollados en operación.

Capítulo 6 Pruebas y Resultados

En este capítulo final se analizan los resultados obtenidos del proyecto MLCOMPET desarrollado, en él se observan los módulos en operación, así como también la información y datos que este proporciona.

6.1 Modulo de búsquedas

Una vez que los productos del usuario han sido obtenidos y se encuentran almacenados en la base de datos, el siguiente paso es obtener los productos de los competidores, para ello se desarrolló el módulo de **Mis Búsquedas** el cual permite al usuario hacer consultas a Mercado Libre con palabras clave (pueden ser nombre del producto, marca, modelo, etc.) con la cual se obtienen todas las publicaciones disponibles relacionadas a la búsqueda realizada (Figura 115).

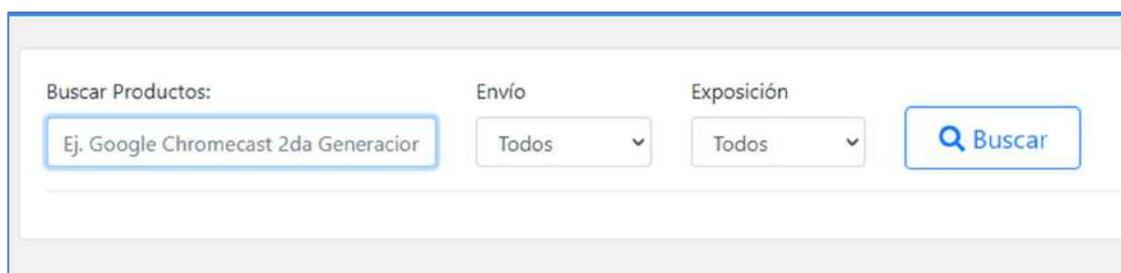
The image shows a search interface with a text input field containing 'Ej. Google Chromecast 2da Generacion', two dropdown menus labeled 'Envío' and 'Exposición' both set to 'Todos', and a 'Buscar' button with a magnifying glass icon.

Figura 115 Módulo de Mis Búsquedas

El módulo incluye filtros que permiten al usuario delimitar la búsqueda por medio de su tipo de envío y el nivel de exposición dentro de la plataforma, con lo que la búsqueda es más precisa. El resultado de esta búsqueda se muestra en forma de listado en la parte inferior de módulo e incluye datos como el nombre del producto, el precio y el vendedor, así también como la cantidad disponible y la cantidad vendida, datos que son utilizados más adelante (Figura 116).

Buscar Productos: Envío Exposición

Guitarra Ibanez Todos Todos

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15

Total : 50 de 713, Página: 1 de 15

<input type="checkbox"/>	Producto	Precio
<input type="checkbox"/>	 Guitarra Electrica Ibanez Grx55b Negra FAMUSICCLUB MLM768377775  Envío Gratis!  Mercado de Pago	Precio: 4399 Tipo Listado: GOLD SPECIAL Disponibles: 1 Vendidos: 2 Envío: DROP OFF
<input type="checkbox"/>	 Guitarra Electrica Ibanez Ps40 Paul Stanley LAFLAUTAMUSICAL MLM768769526  Envío Gratis!  Mercado de Pago	Precio: 7500 Tipo Listado: GOLD PRO Disponibles: 1 Vendidos: 0 Envío: DROP OFF
<input type="checkbox"/>	 Guitarra Electrica Ibanez Grx55b-ye Gio Series Amarilla POLVORA MUSIC MLM783533852  Envío Gratis!  Mercado de Pago	Precio: 5126 Tipo Listado: GOLD PRO Disponibles: 1 Vendidos: 0

Figura 116 Resultado de búsqueda del usuario

En este listado el usuario también puede observar que productos de su catálogo coinciden con la búsqueda realizado, además de la información de las diferentes publicaciones en forma general. Con esta información el usuario puede comparar su producto con el resto de productos disponibles y determinar en base a su experiencia cuál es su competencia (Figura 117).

Catalogo de Productos PROMUSICGEAR:

MLM685629206 - 10 Broches El Cometa Guitarra Eléctrica Alum Varios 843h C \$319.00 (SPECIAL) (DROP_OFF) (Disponibles: 2)

1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23 24

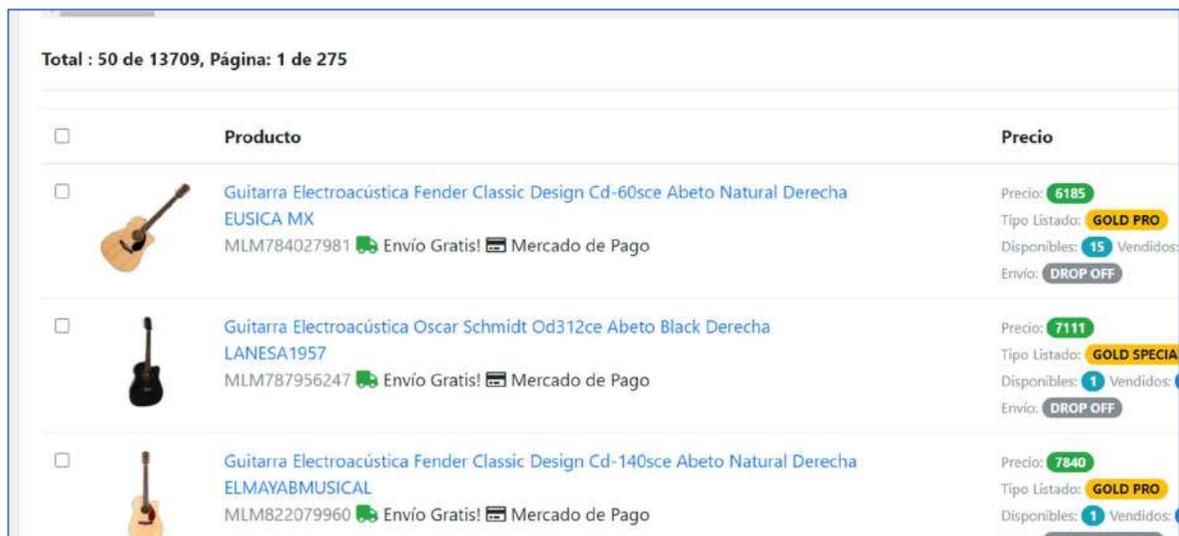
Total : 50 de 136892, Página: 1 de 2738

<input type="checkbox"/>	Producto	Precio
<input type="checkbox"/>	 Guitarra Acustica Curva Nuevos Accesorios Completos Gratis GUITARRAS_PARACHO_MICH MLM620736719  Envío Gratis!  Mercado de Pago	Precio: 1049 Tipo Listado: GOLD PRO Disponibles: 500 Vendidos: 500 Envío: DROP OFF

Figura 117 Vista general de los datos de productos

6.2 Modulo de detección de competencia

Una vez listados los resultados de la búsqueda, el usuario puede analizar los datos de los productos obtenidos, ya que la herramienta le muestra de manera general los datos de las publicaciones de otros vendedores (Figura 118). Entre los datos mostrados tenemos el tipo de publicación, forma de pago, método de envío, precio, cantidad disponible, entre otros.



<input type="checkbox"/>	Producto	Precio
<input type="checkbox"/>	 Guitarra Electroacústica Fender Classic Design Cd-60sce Abeto Natural Derecha EUSICA MX MLM784027981  Envío Gratis!  Mercado de Pago	Precio: 6185 Tipo Listado: GOLD PRO Disponibles: 15 Vendidos: Envío: DROP OFF
<input type="checkbox"/>	 Guitarra Electroacústica Oscar Schmidt Od312ce Abeto Black Derecha LANESA1957 MLM787956247  Envío Gratis!  Mercado de Pago	Precio: 7111 Tipo Listado: GOLD SPECIA Disponibles: 1 Vendidos: Envío: DROP OFF
<input type="checkbox"/>	 Guitarra Electroacústica Fender Classic Design Cd-140sce Abeto Natural Derecha ELMAYABMUSICAL MLM822079960  Envío Gratis!  Mercado de Pago	Precio: 7840 Tipo Listado: GOLD PRO Disponibles: 1 Vendidos: Envío: DROP OFF

Figura 118 Listado de resultados de búsqueda

La aplicación analiza los datos obtenidos de los productos con ayuda del algoritmo de clasificación y los compara con los datos de los productos del usuario para determinar si son competencia, en caso de ser considerados como competencia, el módulo del sistema selecciona aquellos productos habilitando la casilla de verificación que se encuentra del lado izquierdo de la tabla, permitiendo así que el usuario pueda darles seguimiento con el sistema y monitorear su comportamiento (Figura 119).



Figura 119 Detección de competencia

Una vez detectada la competencia por el algoritmo, el usuario puede enfocarse solo en aquellos productos que el sistema ha seleccionado como competidores y descartar aquellos que no, lo cual le ha reducido de manera considerable el tiempo que invertía en realizar este proceso debido a que ya no necesita todos los productos que devuelve la búsqueda.

6.3 Modulo de seguimiento de productos

Una vez seleccionados los productos de la competencia, el usuario puede elegir realizar un seguimiento de los mismos para saber cuándo el otro vendedor haga alguna modificación en la publicación. Para ello el usuario presiona el botón de SEGUIR que se encuentra en la parte superior del listado de los resultados de búsqueda (Figura 120).

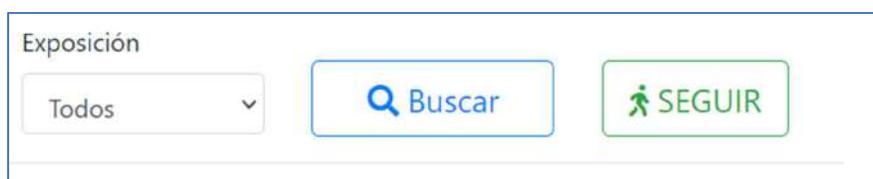


Figura 120 Seguimiento de los productos de la competencia

De este modo los productos seleccionados quedarán guardados y se podrán detectar cambios en las publicaciones, por ejemplo, si llegan a cambiar el precio o cambian de estatus. Para consultar los productos que se están monitoreando se debe acceder a la sección de seguimiento, en donde se muestra un listado de los productos que se están monitoreando actualmente (Figura 121).

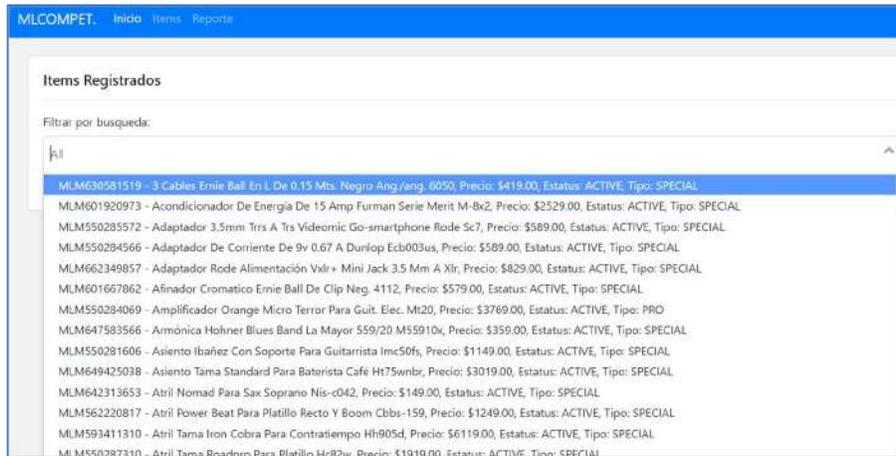


Figura 121 Listado de productos en seguimiento

En esta sección se puede consultar el detalle de un seguimiento, por ejemplo, si el producto del competido ha cambiado su precio o si modifico su estatus, con lo cual el usuario puede determinar si continúa siguiéndolo o lo elimina de la lista. También muestra el precio actual del producto del usuario y el precio más competitivo de los otros vendedores que está siguiendo, permitiendo así modificarlo si lo desea (Figura 122).

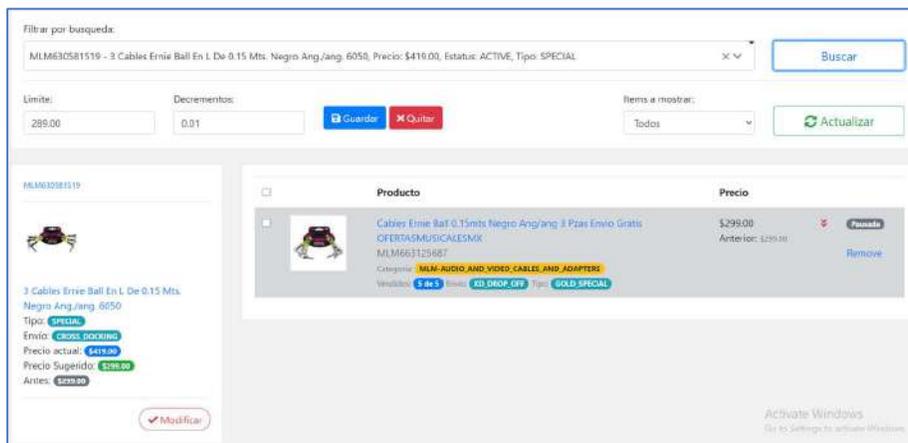


Figura 122 Detalle del seguimiento de la competencia

El seguimiento o monitoreo de los productos de la competencia ha permitido incrementar el número de productos que pueden analizarse, ya que el sistema te notifica solo cuando hay cambios en los productos, por lo que el usuario ya no necesita estar revisando diariamente y solo se enfoca en los que sí han tenido alguna modificación. De acuerdo al análisis realizado se ha incrementado en más de un 50% el número de productos registrados para el seguimiento, lo que se traduce en una mayor eficiencia en el análisis de la competencia.

6.4 Modulo de actualización de precios

El siguiente modulo del proyecto es el de sugerencias y actualización de precios, el sistema analiza los datos de los productos de la competencia y le hace sugerencias de cambio de precio al usuario, tomando como base los mismos datos del catálogo de productos del usuario. En este módulo el usuario puede observar una comparación entre su producto y el producto del competidor con el mejor precio publicado, permitiendo así al usuario actualizar uno o varios registros de modo de tener un precio más competitivo (Figura 123).

Cambios en precios			
<input type="checkbox"/> Producto	Precio	Competencia	Precio
<input type="checkbox"/> Pedal De Efecto Dunlop Mix Reverb M300 [MLM567633116 PROMUSICGEAR2014	Sugerido: \$5059.99 Actual: \$5519.00	Pedal De Efecto Mix Reverb M300 [MLM678973408 FALYMUSIC	\$5090.00
<input type="checkbox"/> Microfono Condensador Rode Nt Usb [MLM550282176 PROMUSICGEAR2014	Sugerido: \$3889.00 Actual: \$4719.00	Rode Microfono Podcast Nt-usb [MLM620986221 TECNOPLANET MEXICO	\$3899.00
<input type="checkbox"/> Parche Remo Powersonic De 22 Transparente Pw-1322-00 [MLM553327684 PROMUSICGEAR2014	Sugerido: \$1148.99 Actual: \$1629.00	Remo Pw132200 Parche Powersonic De 22 Transparente Bombo [MLM630441337 BACKSTAGE_MTY	\$1190.00
<input type="checkbox"/> Combo Ibanez Para Bajo Eléctrico 10w 1 X 6.5 lbz10b-n [MLM550287947 PROMUSICGEAR2014	Sugerido: \$1818.99 Actual: \$1889.00	Amplificador P/bajo Electrico Ibanez 10w [MLM639568110 THE_GUITAR_HOUSE_MEX	\$1819.00

Figura 123 Sugerencias y Cambio de Precio en productos

En la imagen anterior pueden observarse aquellos registros en los que el usuario necesita verificar los precios de sus productos, debido a que hay productos de otros vendedores con precios más competitivos, esto ayuda al usuario a realizar un análisis de sus costos y precios, de modo que puede optar por hacer un cambio en el precio publicado de modo de volverlo más competitivo que el de otros vendedores (Figura 124).



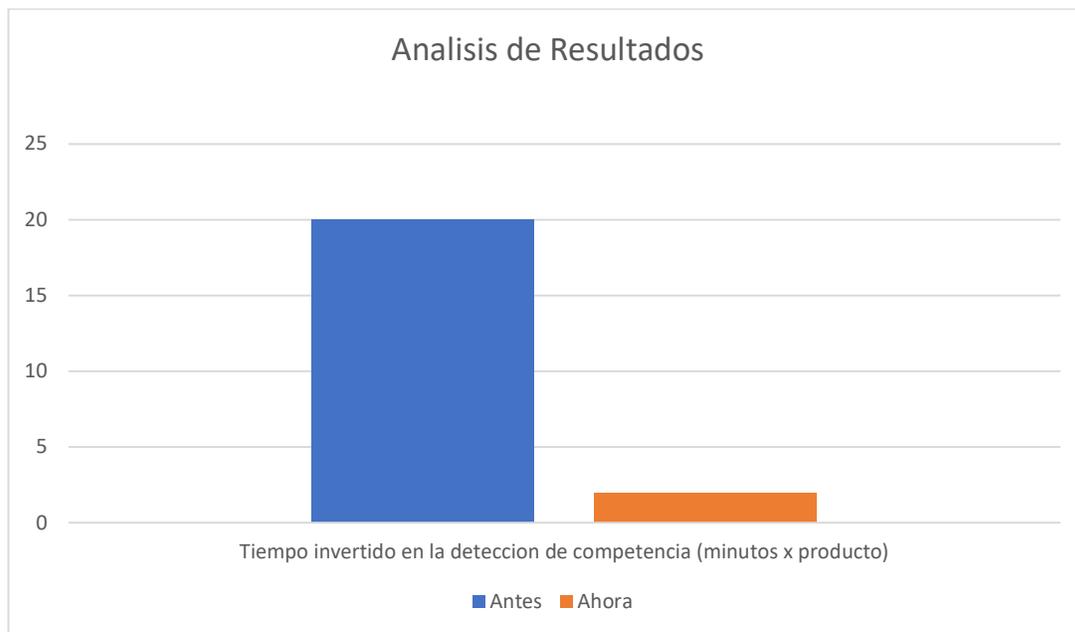
<input type="checkbox"/>	Producto	Precio
<input checked="" type="checkbox"/>	Pedal De Efecto Dunlop Mxr Reverb M300 [MLM567633116] PROMUSICGEAR2014	Sugerido: \$5059.99 Actual: \$5519.00
<input checked="" type="checkbox"/>	Micrófono Condensador Rode Nt Usb [MLM550282176] PROMUSICGEAR2014	Sugerido: \$3889.00 Actual: \$4719.00
<input checked="" type="checkbox"/>	Parche Remo Powersonic De 22 Transparente Pw-1322-00 [MLM553327684] PROMUSICGEAR2014	Sugerido: \$1148.99 Actual: \$1629.00
<input type="checkbox"/>	Combo Ibañez Para Bajo Eléctrico 10w 1 X 6.5 lbz10b-n [MLM550287947] PROMUSICGEAR2014	Sugerido: \$1818.99 Actual: \$1889.00

Figura 124 Actualización de precios en los productos seleccionados

El usuario puede seleccionar y actualizar los precios de 1 o varios productos al mismo tiempo, esto ha permitido al usuario poder responder de manera más rápida al comportamiento de su competencia, logrando así mantener posicionados sus productos dentro de la plataforma de ventas.

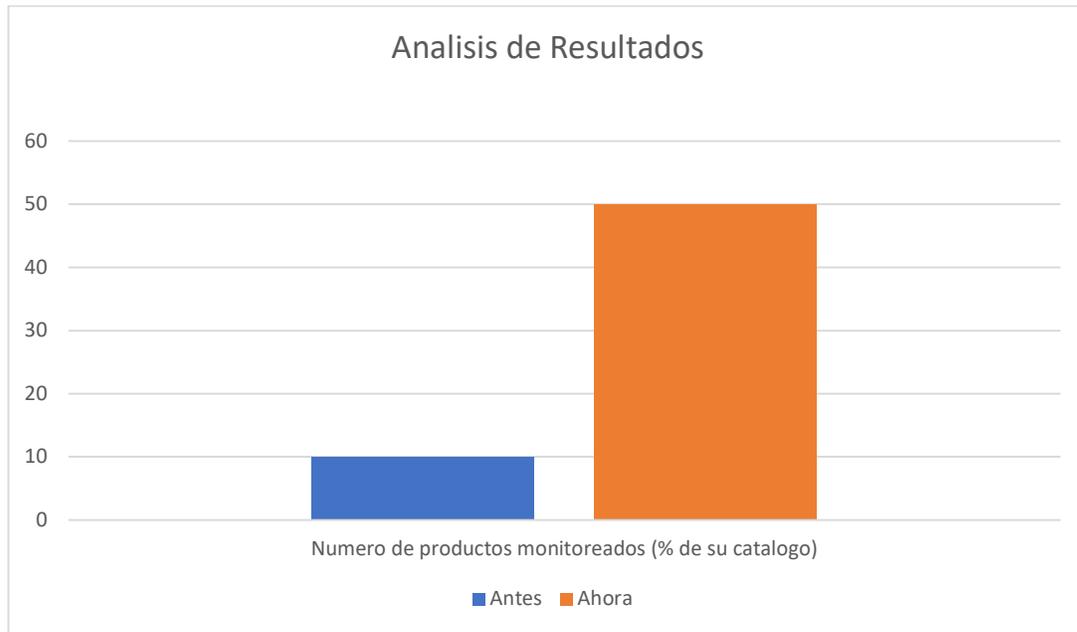
6.5 Análisis de Resultados

A continuación, se listan de manera gráfica los resultados obtenidos posterior a la implementación del proyecto en la empresa. En la Grafica 1 se puede observar una reducción considerable en el tiempo invertido para realizar la detección de competencia, donde se redujo cerca del 100% pasando de 20 minutos a tan solo 2 minutos por cada producto aproximadamente.



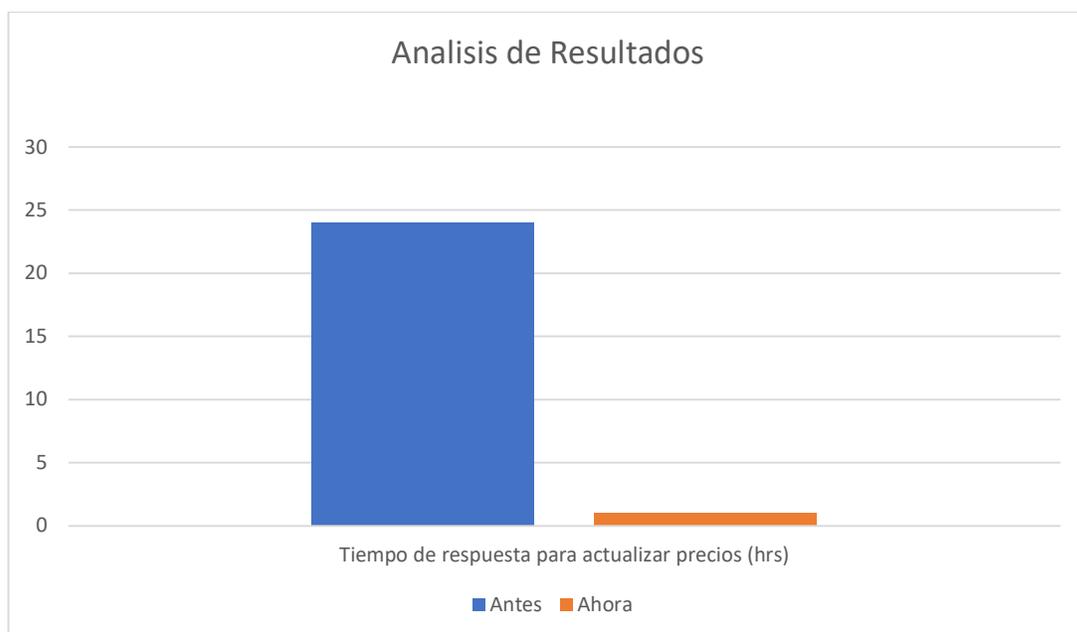
Grafica 1 Resultados de tiempo invertido en la detección de competencia

En la siguiente grafica (Grafica 2) se puede observar el incremento en el número de los productos que se están monitoreando, anteriormente solo se podía llegar a monitorear cerca del 10% de los productos de su catálogo, mientras que ahora se ha podido llegar cerca del 50% de los productos monitoreados.



Grafica 2 Resultados de numero de productos monitoreados

Por último, la siguiente grafica se presentan los resultados relacionados al tiempo de respuesta del usuario a cambios en las publicaciones de otros vendedores, en donde podemos observar que en lugar de esperar 1 día aproximadamente a realizar nuevamente el análisis y detectar si hubo cambios en los productos de sus competidores, ahora puede responder prácticamente de inmediato ya que el sistema le notifica de los cambios y puede actualizar el precio inmediatamente.



Grafica 3 Resultados del tiempo de respuesta para actualizar precios

Conclusiones y trabajos futuros

Una vez concluido el proyecto MLCOMPET para esta primera versión y después de haber analizado los resultados obtenidos, se ha podido llegar a la conclusión de que al implementar una herramienta de software como la desarrollada en este proyecto, marca en gran medida la diferencia en la forma de hacer las cosas, ya que se ha demostrado mediante resultados como pueden beneficiarse las empresas al optimizar sus procesos administrativos como lo es el análisis de la competencia, el cual es de gran importancia debido a la gran cantidad de participante en el mercado actual, las empresas deben de contar con las herramientas necesarias que logren diferenciarlos de los demás.

El uso de algoritmos de inteligencia artificial aplicados al desarrollo de herramientas que permitan facilitar el trabajo de las empresas y hacerlas a su vez más productivas, abre un abanico grande de posibilidades que se pueden implementar para optimizar los diferentes procesos dentro de una empresa, no solo en el área de ventas, si no también, en el área financiera, operativa, recursos humanos entre muchas otras. Cada vez más, la inteligencia artificial va ganando presencia en desarrollo de diversas aplicaciones, por lo que se puede esperar que, en un futuro no muy lejano, se vea de manera natural su presencia en la construcción de proyectos de software de diversa índole y que su uso se extienda cada vez más a más personas, lo que podría favorecer su investigación y crecimiento.

El proyecto MLCOMPET se desarrolló con la finalidad de resolver la problemática relacionada a la detección y análisis de la competencia, para publicaciones de productos en Mercado Libre, sin embargo su desarrollo aun puede ampliarse a otras áreas de la misma plataforma, por ejemplo, análisis de ventas,

implementación de lenguaje natural para la implementación de respuestas automáticas, integración con otras plataformas como Amazon o Linio, o incluso sistemas administrativos como ERP los cuales almacenas datos de todas las áreas de una empresa, datos de los cuales se puede obtener información importante para la toma de decisiones de la empresa.

Por último, me gustaría agregar que el desarrollo de este proyecto ha permitido explorar algunos algoritmos de inteligencia artificial mas complejos, los cuales se podrían utilizar de forma individual o en conjunto con el fin de obtener resultados cada vez mejores que permitan que la herramienta pueda seguir desarrollándose.

Bibliografía

- APD, R. (2019). *apd*. Obtenido de Machine Learning: <https://www.apd.es/algoritmos-del-machine-learning/>
- Changhui. (2016). Sale Decision Analysis System Design Based on Data Warehouse. *IEEE*, 1.
- COFECE. (2018). Obtenido de <https://www.cofece.mx/wp-content/uploads/2018/05/1lacompetenciaeconom.pdf>
- ComScore, & EMarketer. (Mayo de 2018). *Statista*. Obtenido de <https://www.statista.com/statistics/321543/latin-america-online-retailer-visitors/>
- CreceNegocios. (10 de 10 de 2019). *CreceNegocios*. Obtenido de Analisis de la competencia: <https://www.crecenegocios.com/analisis-de-la-competencia/>
- DataScience*. (28 de Marzo de 2018). Obtenido de Data Science: <http://datascience.esy.es/wiki/preparacion-de-datos/>
- Dr. Shahid & Kansana, K. &. (2016). A Review Paper on E-Commerce. *TIMS 2016-International Conference*. Gwalior: https://www.researchgate.net/publication/304703920_A_Review_Paper_on_E-Commerce.
- Endeavor. (2010). *Emprendedor*. Obtenido de La competencia directa e indirecta: <http://www2.esmas.com/emprendedor/herramientas-y-articulos/marketing/184455/competencia-competencia-directa-competencia-indirecta/>
- ENDUTIH. (2018). Obtenido de https://www.inegi.org.mx/contenidos/saladeprensa/boletines/2019/OtrTemEcon/ENDUTIH_2018.pdf
- Forbes. (31 de Enero de 2018). *Pymes Mexicanas, Un panorama para 2018*. Obtenido de <https://www.forbes.com.mx/pymes-mexicanas-un-panorama-para-2018/>
- Han, J. (2011). *Data Mining: Concepts and Techniques, 3rd Ed*. Morgan Kaufmann.
- Kamplán, A., & Haenlein, M. (2019). *Siri, Siri in my Hand, who's the Fairest in the Land? On the Interpretations, Illustrations and Implications of Artificial Intelligence*. Elsevier: Business Horizons, 62(1), 15-25.
- Khan, A. G. (2016). Electronic Commerce: A Study on Benefits and Challenges in an Emerging Economy. *Global Journal of Management and Business Research: B Economics and Commerce*.
- Linio. (2018). *Índice Mundial de Comercio Electrónico*. Obtenido de <https://www.linio.com.mx/sp/indice-ecommerce>
- Maimon, O., & Rokach, L. (2010). *Data Mining and Knowledge Discovery Handbook*. New York: Springer.
- Mckinsey, & Company. (2018). *Skill Shift Automation and the future of the work force*. Obtenido de <https://www.mckinsey.com/~media/mckinsey/featured%20insights/future%20of%20organizations/skill%20shift%20automation%20and%20the%20future%20of%20the%20workforce/mgi-skill-shift-automation-and-future-of-the-workforce-may-2018.ashx>

- Mercado Libre Developers*. (04 de 08 de 2020). Obtenido de https://developers.mercadolibre.com.ar/es_ar/categorias-y-atributos
- MercadoLibre. (2018). Obtenido de <https://www.marketplacepulse.com/stats/mercadolibre/mercado-libre-number-of-sellers-81>
- Microsoft. (01 de 08 de 2019). *Microsoft*. Obtenido de Data Mining: <https://docs.microsoft.com/es-es/analysis-services/data-mining/data-mining-ssas?view=sql-server-2017>
- Mozina, M., Demsar, J., Kattan, M., & Zupan, B. (2004). Nomograms for Visualization of Naive Bayesian Classifier. *Springer, Berlin, Heidelberg*.
- Pirani, Z. (2017). Analysis and Optimization of Online Sales of Products. *IEEE*, 1.
- Poole, & David. (2018). *Computational Intelligence: A Logical Approach*. Nueva York: Oxford University.
- Radoslav Fasuga, P. S. (2014). The Method of Automated Monitoring of Product Prices and Market Position in Relation to Competition Quotes. *IEEE*, 1.
- Russell, J., S., & Norvig. (2009). *Artificial intelligence: a modern approach (3a Ed.)*. Upper Saddle River, N.J.: Prentice Hall.
- Russell, S., & Norvig, P. (2009). *Inteligencia Artificial: Un Enfoque Moderno*. Prentice Hall.
- Schank, & C., R. (1991). Where's the AI. *AI Magazine*, 12(4): 38.
- ScikitLearn*. (s.f.). Obtenido de ScikitLearn: https://scikit-learn.org/stable/modules/naive_bayes.html
- Sharma Mayank, V. S. (2016). Enhancing sales by using information with the help of object detection and object storage. *IEEE*.
- Sutton, R. S., & Barto, A. G. (1998). *Reinforcement Learning: An Introduction*.
- Toptal. (s.f.). *20 IU de comercio electrónico que facilitan las compras*. Obtenido de <https://www.toptal.com/designers/mobile/20-ecommerce-uis-that-make-shopping-easier>
- WikiSaber. (s.f.). *El comercio electrónico sigue aumentando su tendencia*. Obtenido de <https://www.wikisaber.es/comercio-electronico-sigue-aumentando-su-tendencia/>