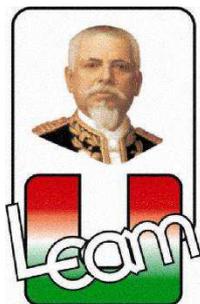


UNIVERSIDAD LAICA “ELOY ALFARO” DE MANABÍ

FACULTAD DE CIENCIAS INFORMÁTICAS



TESIS DE GRADO

Previo a la obtención del título de:

INGENIERA EN SISTEMAS

TEMA:

ESTUDIO DEL ANÁLISIS PREDICTIVO PARA EL DESARROLLO DE APLICACIONES EN LA FACULTAD DE CIENCIAS INFORMÁTICAS DE LA ULEAM, A TRAVÉS DE LA PLATAFORMA DE COMPUTACIÓN EN LA NUBE MICROSOFT AZURE.

AUTORA:

Verónica Estefanía Muentes Molina

DIRECTORA DE TESIS:

Lic. Dolores Muñoz Verduga, Mg.

Manta - Manabí - Ecuador

2015

CERTIFICACIÓN

En mi calidad de Directora de Tesis de Grado de la Facultad de Ciencias Informáticas de la Universidad Laica “Eloy Alfaro” de Manabí, certifico haber dirigido y revisado la tesis sobre el tema: “ESTUDIO DEL ANÁLISIS PREDICTIVO PARA EL DESARROLLO DE APLICACIONES EN LA FACULTAD DE CIENCIAS INFORMÁTICAS DE LA ULEAM, A TRAVÉS DE LA PLATAFORMA DE COMPUTACIÓN EN LA NUBE MICROSOFT AZURE” de la estudiante Verónica Estefanía Muentes Molina con C.I. 131393806-8, considero que el mencionado trabajo investigativo reúne los requisitos y méritos suficiente para ser sometido a la evaluación del jurado examinador que las autoridades designen.

En honor a la verdad.

Lic. Dolores Muñoz Verduga, Mg.
DIRECTORA DE TESIS

**TRIBUNAL DE GRADUACIÓN
DECLARACIÓN EXPRESA**

DEDICATORIA

“No basta dar pasos que un día puedan conducir hasta la meta, sino que cada paso ha de ser una meta, sin dejar de ser un paso”

Johann P. Eckermann.

Dedico este trabajo a mi familia, por su confianza, amor, apoyo incondicional y por enseñarme cada día el valor de la perseverancia. A mi directora por la ayuda incondicional, la dedicación y paciencia a lo largo del trayecto, y a mis familiares, profesores y amigos que han estado ahí durante mi vida profesional y personal, por sus consejos y enseñanzas compartidas.

Verónica Muentes

AGRADECIMIENTO

“El agradecimiento es la memoria del corazón”

Lao-Tsé

Agradezco a todas las personas que me ayudaron a lo largo del camino, y a todas las personas que de una o cierta forma hicieron posible este objetivo. Gracias infinitas.

RESUMEN

El presente proyecto de tesis parte de la inteligencia artificial y su facultad para lograr que una máquina aprenda, y se manifieste por medio del Machine Learning, razón de mi estudio, que a través del Cloud Computing, por medio de la plataforma Microsoft Azure demuestra lo fácil que es, y porque se está convirtiendo en un fenómeno mundial que no se puede dejar de explorar.

El motivo principal del estudio realizado, se encuentra en el nivel de desconocimiento existente por los alumnos de últimos niveles de la Carrera de Ingeniería de Sistemas de la Facultad de Ciencias Informáticas de la Universidad Laica “Eloy Alfaro” de Manabí, futuros profesionales y potenciales expertos en predicciones basadas en Aprendizaje automático con Azure o con otra plataforma cloud.

Antes las empresas necesitaban recursos económicos elevados, mucho tiempos para estudios, algoritmos complejos, pero gracias al Cloud computing, se pueden obtener los mismos y mejores resultados con menos tiempo y dinero.

Los estudiantes deben ser una fuente de desarrollo y conocimiento, los resultados que se obtuvieron determinaron la exposición del funcionamiento de Machine Learning en Azure, promoviéndolo por medio de conceptos y ejemplos que se materializaron por medio de una guía. Esta reunió las características más importantes del Machine Learning desde una perspectiva cloud, su utilización y conocimiento, les permitió tener mayores ideas sobre cómo potenciar y dar valor agregado a sus proyectos, así como la otorgación de experiencia y mayores retos, con las facilidades que dan las buenas y nuevas tecnologías. Como resultado además se obtuvo un recurso para el aprendizaje y uso de esta rama de la inteligencia artificial, para el público en general.

ÍNDICE

CERTIFICACIÓN	2
TRIBUNAL DE GRADUACIÓN	3
DECLARACIÓN EXPRESA	3
DEDICATORIA	3
AGRADECIMIENTO	4
RESUMEN	5
NDICE DE IMÁGENES	9
ÍNDICE DE TABLAS	12
INTRODUCCIÓN	13
UBICACIÓN Y CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN	17
PLANTEAMIENTO DEL PROBLEMA	18
DEFINICIÓN DEL PROBLEMA	19
DELIMITACIÓN DEL CONTENIDO	19
DELIMITACIÓN ESPACIAL	19
DELIMITACIÓN TEMPORAL	19
OBJETIVOS	20
OBJETIVO GENERAL	20
OBJETIVOS ESPECÍFICOS DE INVESTIGACIÓN	20
OBJETIVOS ESPECÍFICOS DE RESOLUCIÓN DEL PROBLEMA	20
JUSTIFICACIÓN	21
IMPACTOS ESPERADOS	23
IMPACTO TECNOLÓGICO	23
IMPACTO SOCIAL	23
CAPITULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN	24
1.1 INTRODUCCIÓN	24
1.2. ANTECEDENTES DE INVESTIGACIONES RELACIONADAS AL TEMA.	24
INVESTIGACIONES ACERCA DE MICROSOFT AZURE Y CLOUD COMPUTING.	24
INVESTIGACIONES ACERCA DE MACHINE LEARNING.	26
1.3 DEFINICIONES CONCEPTUALES	29
1.3.1 Aprendizaje	29

1.3.2	Aprendizaje de Máquina - Machine Learning	31
1.3.3	Tecnología	33
1.3.4	Tecnología de la información	34
1.3.5	Arquitectura cliente/Servidor	40
1.3.6	Aplicaciones Informáticas	48
1.3.7	Cloud Computing – Computación en la nube	51
1.3.8	Arquitectura de un Web Service	62
1.3.9	Infraestructura de TI	64
1.4	FUNDAMENTACIÓN LEGAL	68
1.5	CONCLUSIONES RELACIONADAS AL MARCO TEÓRICO EN REFERENCIA AL TEMA DE INVESTIGACIÓN	69
CAPITULO II: DIAGNÓSTICO O ESTUDIO DE CAMPO		70
2.1	INTRODUCCIÓN	70
2.2	TIPO DE INVESTIGACIÓN	70
2.3	MÉTODOS DE INVESTIGACIÓN	71
2.4	HERRAMIENTA(S) DE RECOLECCIÓN DE DATOS	72
2.5	FUENTES DE INFORMACIÓN DE DATOS	73
2.6	INSTRUMENTAL OPERACIONAL	74
2.6.1	Estructura y características de lo(s) instrumento(s) de recolección de datos	74
2.7	ESTRATEGIA OPERACIONAL PARA LA RECOLECCIÓN Y TABULACIÓN DE DATOS	75
2.7.1	Plan de recolección, tabulación, análisis e interpretación de los datos	75
2.8	PLAN DE MUESTREO	76
2.8.1	Segmentación - técnica de muestreo - tamaño de la muestra	76
2.9	PRESENTACIÓN Y ANÁLISIS DE LOS RESULTADOS	77
2.9.1	Presentación y descripción de los resultados obtenidos	77
2.9.2	Informe final del análisis de los resultados	102
CAPITULO III: DISEÑO DE LA PROPUESTA		104
3.1	INTRODUCCIÓN	104
3.2	TITULO DE LA PROPUESTA	104
3.3	DESCRIPCIÓN DE LA PROPUESTA	104

3.4. JUSTIFICACIÓN DE LA PROPUESTA	105
3.5.1 Objetivo General	106
3.5.2 Objetivos Específicos	106
3.6 PROPUESTA	106
Introducción.	106
Computación en la Nube	107
Microsoft Azure	109
Machine Learning	109
¿Aprendizaje supervisado? ó ¿Aprendizaje no supervisado?	110
Pero, ¿Qué hace Microsoft Azure con Machine Learning?	114
Ventajas y Desventajas de Microsoft Azure Machine Learning	115
Machine Learning y Minería de Datos.	116
Características para determinar si tu solución debe ser Machine Learning. 117	
Pasos para construir una solución Machine Learning en Microsoft Azure ...118	
Estudio Machine Learning.	119
Introducción al Estudio ML	120
Práctica en el Studio ML	122
¿Qué puedes hacer con ML?	137
Empresas que usan ML en Azure	138
Costos por uso del Studio Machine Learning	139
CAPITULO IV: EVALUACIÓN DE RESULTADOS	141
4.1 INTRODUCCIÓN	141
4.2 SEGUIMIENTO Y MONITOREO DE RESULTADOS	141
CONCLUSIONES	145
RECOMENDACIONES	146
ANEXO 1. PLANTEAMIENTO DEL PROBLEMA	158
ANEXO 2. Encuesta No. 1	159
ANEXO 3. Encuesta No. 2	162
ANEXO 4. Encuesta No. 3	164
ANEXO 5. Encuesta No. 4	166
GLOSARIO	168

ÍNDICE DE IMÁGENES

Ilustración 1 Arquitectura cliente-servidor	40
Ilustración 2 Arquitectura web services	63
Ilustración 3 Autenticación y Autorización	66
Ilustración 4 Plan de muestreo	76
Ilustración 5 Encuesta 1 - Pregunta 1	78
Ilustración 6 Encuesta 1 - Pregunta 2	79
Ilustración 7 Encuesta 1 - Pregunta 3	80
Ilustración 8 Encuesta 1 - Pregunta 4	81
Ilustración 9 Encuesta 1 - Pregunta 5	82
Ilustración 10 Encuesta 1 - Pregunta 6	83
Ilustración 11 Encuesta 1 - Pregunta 7	84
Ilustración 12 Encuesta 1 - Pregunta 8	85
Ilustración 13 Encuesta 1 - Pregunta 9	86
Ilustración 14 Encuesta 1 - Pregunta 10	87
Ilustración 15 Encuesta 2 – Pregunta 1	88
Ilustración 16 Encuesta 2 - Pregunta 2	89
Ilustración 17 Encuesta 2 - Pregunta 3	90
Ilustración 18 Encuesta 2 - Pregunta 4	91
Ilustración 19 Encuesta 2 - Pregunta 5	92
Ilustración 20 Encuesta 2 - Pregunta 6	93
Ilustración 21 Encuesta 3 - Pregunta 1	94
Ilustración 22 Encuesta 3 - Pregunta2	95
Ilustración 23 Encuesta 3 - Pregunta3	96
Ilustración 24 Encuesta 3 - Pregunta4	97
Ilustración 25 Encuesta 3 - Pregunta5	98
Ilustración 26 Encuesta 3 - Pregunta 6	99
Ilustración 27 Encuesta 3 - Pregunta 7	100
Ilustración 28 Encuesta 3 - Pregunta 8	101
Ilustración 29 Alcance de Machine Learning	107
Ilustración 30 Estructura de la computación en la nube	107
Ilustración 31 Amazon Web Services	108
Ilustración 32 IBM Smart Cloud	108
Ilustración 33 Microsoft Azure	108
Ilustración 34 Google Cloud Platform	108
Ilustración 35 ¿Qué es Machine Learning?	109
Ilustración 36 Características de Machine Learning	110
Ilustración 37 Modelo de predicción de valor	111
Ilustración 38 Asociación	111
Ilustración 39 Regresión lineal	112
Ilustración 40 gradiente descendente	113

Ilustración 41 Machine Learning	114
Ilustración 42Ventajas Machine Learning	115
Ilustración 43 Desventajas Machine Learning	115
Ilustración 44 Minería de datos vs Machine Learning	116
Ilustración 45 Servidores de computadoras	117
Ilustración 46 Necesidades Machine Learning	117
Ilustración 47 Pasos para construir una solución Machine Learning	118
Ilustración 48 Estudio Machine Learning	119
Ilustración 49 ML Estudio	120
Ilustración 50 Opciones ML	120
Ilustración 51 Nuevo ML	121
Ilustración 52 Componentes	121
Ilustración 53 Opciones 2	121
Ilustración 54 Características auto	122
Ilustración 55 Iniciar sesión	122
Ilustración 56 Abrir en Studio	123
Ilustración 57 Probar Machine Learning	124
Ilustración 58 ML Studio	124
Ilustración 59 Nuevo ML	124
Ilustración 60 Cargar Dataset	125
Ilustración 61 Cargar Módulo	125
Ilustración 62 Experimento	126
Ilustración 63 Nuevo experimento	126
Ilustración 64 Primer experimento	127
Ilustración 65 Selección de DataSet	127
Ilustración 66 Opciones DataSet	127
Ilustración 67 Proyectar columnas	128
Ilustración 68 Excluir campos	128
Ilustración 69 Sustitución de valores nulos	129
Ilustración 70 Valores nulos	129
Ilustración 71 Ejecución de experimento	130
Ilustración 72 Seleccionar columnas	130
Ilustración 73 División	131
Ilustración 74 Modelo de entrenamiento	131
Ilustración 75 Modelo de puntaje	132
Ilustración 76 Campo de entrenamiento	132
Ilustración 77 Correr modelo	132
Ilustración 78 Visualización de datos	133
Ilustración 79 Ejecución ML	134
Ilustración 80 Web Services	135
Ilustración 81 Respuesta Web Services	135
Ilustración 82 Web Services de experimento	136

Ilustración 83 Api Key.....	136
Ilustración 84 Respuesta de solicitud.....	136
Ilustración 85 Solicitud de Web Services.....	137
Ilustración 86 Análisis.....	137
Ilustración 87 Búsqueda.....	137
Ilustración 88 Buscadores.....	138
Ilustración 89 MAX 451.....	138
Ilustración 90 OSisoft.....	139
Ilustración 91 ThyssenKrupp.....	139
Ilustración 92 Theorem.....	139
Ilustración 93 Costos Azure.....	140
Ilustración 94 Resultados antes del estudio.....	144
Ilustración 95 Resultados después del estudio.....	144

ÍNDICE DE TABLAS

Tabla 1	Indicador de situación	143
---------	------------------------------	-----

INTRODUCCIÓN

El campo de la Inteligencia artificial sigue dando de qué hablar a pesar de sus años, y es que es impresionante la forma en la que se encarga de imitar a una persona, y no hablamos de diferentes partes del cuerpo, sino del cerebro específicamente. Lo más complejo que existe es el cerebro humano, y que una ciencia haya logrado crear algo parecido es un gran logro. Aunque ahora ya no es tan sorprendente, su existencia en la antigüedad no se pensaba fuera posible, y no es más que años y años de desarrollo e investigación.

Y es que con la inteligencia artificial se abrieron varias ramas de conocimiento debido a la amplitud por obvias razones, una de ellas fue el Machine Learning o por su nombre en español, Aprendizaje Automático. Campo que aunque ha sido utilizado desde su creación, siempre ha denotado una amplia complejidad; razón por la que su ejecución siempre ha sido limitada por el tiempo, dinero y conocimiento. Ésta busca en datos pasados o existentes, por medio de patrones, la construcción de modelos que puedan predecir valores que ayuden a la toma de decisiones.

El ML está dividido por dos amplios aprendizajes, el supervisado y el no supervisado.

En el aprendizaje supervisado examina predicciones en cuanto a valores numéricos, y se divide por dos tipos de modelos, el de regresión que trabaja con valores continuos como pronósticos de precios de bicicletas el año que viene, y el de clasificación que lo hace con valores discretos, muy usado en la medicina por pronosticar el cáncer o la diabetes.

En el aprendizaje no supervisado por el contrario, sondea predicciones de valores tipo cadena, y se divide en clustering, modelo que busca agrupar datos según características como la edad, el sexo, y asociación que sondea

características y las va asociando con algo, por ejemplo el consorcio de la menstruación con el sexo femenino, la edad, y los cólicos.

Son diferentes formas de ver el aprendizaje, su mayor diferencia es que en el aprendizaje supervisado conoces el valor a predecir, pero el aprendizaje no supervisado saca la parte más inteligente de la máquina porque aprenderá de su propio conocimiento e irá ordenando la información de tal forma que llegue a una conclusión.

El Machine Learning se ha encargado de hacer inteligentes a las máquinas a lo largo del tiempo, pero con un alto costo.

La tecnología ha seguido avanzando y con ella la forma en que se analizan los datos. Se han dado pasos gigantes en cuanto a Data Science (Ciencia de Datos), razón por la que grandes tecnologías, como lo es Big Data gestiona inmensas cantidades de información desde gigas hasta exas, haciendo lo que una base de datos relacional no puede; esto en cuanto a data, en cuanto a infraestructura, software y demás podemos encontrar al, rey de reyes, Cloud Computing que facilitó la creación de las famosas plataformas de servicios o cloud, en las que se tiene el software como servicio, infraestructura como servicio, y hasta plataforma como servicio. Esto ha abierto una gran brecha entre el antes y el ahora, ahora tenemos la posibilidad de tener más pagando menos con su uso.

Las plataformas cloud han abierto puertas y posibilidades para todos, las implementaciones representaban altos costos, comprar el hardware, el software, luego configurarlos costaba un dineral, sin contar cuando la infraestructura se quedaba corta con el nuevo software, y es que mientras exista la programación siempre van a haber nuevas actualizaciones; era una cosa sólo para empresas millonarias y más. Todo cambió.

Entre las plataformas más conocidas y mejores en la actualidad se encuentran Microsoft Azure y Amazon Web Services. Pero si se debe escoger

una, la inclinación es hacia Azure por sus interfaces intuitivas y sus controles de confidencialidad y seguridad que no tiene ninguna otra. Esta plataforma ha tenido un crecimiento tremendo en estos dos últimos años, tiene un sin número de servicios, tiene certificaciones ISO, sin mencionar que está reconocida por Gartner, una de las mayores empresas consultoras y de investigación tecnológica a nivel mundial, como la mejor plataforma en ofrecer infraestructura como servicio.

El aprendizaje automático es uno de los servicios más nuevos que tiene ahora Azure. Es sin duda una herramienta de primera, llamada Studio Machine Learning, en español Estudio de Aprendizaje Automático, ésta permite crear un modelo por medio de un gráfico de árbol compuesto por componentes para su predicción, en el que se arrastra y se coloca cada uno. Permite la reutilización de componentes, la retroalimentación de datos locales o alojados en la nube, el trabajo en equipo, la publicación de modelos por medio de servicios webs y la ingeniería de componentes.

Ahora es todo más fácil y con menor costo. Alumnos, profesores, emprendedores, empresas pequeñas, medianas, grandes, pueden darle uso. No existen impedimentos para no hacer a nuestras aplicaciones o implementaciones más inteligentes, con una solución que se adapta a nuestro presupuesto, porque el costo es proporcional al uso y al tiempo de ejecución; y es gratuita si sólo queremos aprender.

Entonces, ¿Por qué no seguir avanzando e innovando?, Es verdaderamente útil dar mayor acceso al conocimiento de aprendizaje automático por medio de guías y libros dirigidos para ingenieros, no sólo para físicos-matemáticos, textos en español y no sólo en inglés, para que esté disponible para todos; sin olvidar que también es necesario que no sólo haya interés de estudiarlo porque represente dinero, sino hacer conciencia de que el mundo en que vivimos sólo llegará lejos si se siguen haciendo nuevas investigaciones.

La actual existencia de máquinas inteligentes y el estudio de comportamientos como la predicción del cáncer, son sólo pequeñas cosas comparadas con las que se vienen; sin contar que después de diez años probablemente haya toda una población de artefactos, dispositivos y máquinas inteligentes.

UBICACIÓN Y CONTEXTUALIZACIÓN DE LA INVESTIGACIÓN

Ubicación:

Ciudadela Universitaria, Calle 12 Vía San Mateo.

Manta - Manabí - Ecuador

Contextualización:

El estudio demostró que por medio del modelo “Linear Regression” aplicado en Machine Learning de Microsoft Azure, causó un gran impacto en los universitarios de la Carrera de Ingeniería de Sistemas, promoviendo la investigación y el aprendizaje en su más alto nivel; así como aportando valor a sus aplicaciones web y experiencia requerida en el ámbito laboral.

PLANTEAMIENTO DEL PROBLEMA

Desde la existencia de la inteligencia artificial, el Machine Learning ha ido perfeccionando sus predicciones, pero había sido ignorado por sus altos costos y el tiempo que requería el diseñar un modelo de predicción.

Actualmente existen un sin número de herramientas informáticas de fácil acceso, hacer aplicaciones web, predecir valores futuros, manejar servidores, o configurar una red es algo que gracias al Cloud Computing es más fácil y rápido, permitiendo el uso de cualquier tipo de usuario, desde principiantes hasta avanzados.

Existen diversas plataformas cloud, desde IBM SmartCloud, hasta Microsoft Azure, si se sabe sacar el potencial puede dar resultados óptimos, pero es pequeño el público que conoce, desde estudiantes hasta empresas. Se sigue trabajando y estudiando de forma convencional e innovar es necesario. A nivel mundial en la rama de informática se habla del cloud computing y de las distintas plataformas que prestan servicios en la nube. Studio ML es un servicio paas, una idea de laboratorio científico informático predictivo, componentes para arrastrar y colocar, listo para hacer gráficos de árbol que permitan predicciones fáciles de entrenar, una galería de modelos para usar y Web Services a la mano para publicar el modelo y consumir.

DEFINICIÓN DEL PROBLEMA

DELIMITACIÓN DEL CONTENIDO.

Campo: Informática.

Áreas: Inteligencia Artificial.

Aspecto: Valoración de la necesidad de un estudiante de 7mo, 8vo y 9no nivel en cuanto al aprendizaje de herramientas de Cloud Computing (Plataforma de Servicios: Microsoft Azure) y la generación de interés por utilizar y conocer Machine Learning en Azure.

DELIMITACIÓN ESPACIAL.

Esta investigación se realizó con los estudiantes de los últimos niveles (séptimo, octavo y noveno) de la Facultad de Ciencias Informáticas de la Universidad Laica “Eloy Alfaro” de Manabí.

DELIMITACIÓN TEMPORAL.

El estudio se realizó en el período comprendido entre el 12 de Noviembre del año 2014 y el 30 de Junio del 2015.

OBJETIVOS

OBJETIVO GENERAL

Analizar e investigar el uso del modelo de Regresión Lineal en Machine Learning de Microsoft Azure, orientado a los estudiantes de últimos niveles (séptimo, octavo y noveno) de la Facultad de Ciencias Informáticas de la ULEAM, para dar valor a las aplicaciones web, con el fin de innovar metodologías, promover el Cloud Computing, aumentar el aprendizaje y dotar de mayor experiencia.

OBJETIVOS ESPECÍFICOS DE INVESTIGACIÓN

- Definir el nivel de conocimiento, interés y uso de esta tecnología en los estudiantes de últimos niveles.
- Analizar el funcionamiento básico del modelo de Regresión Lineal usado en Machine Learning.
- Realizar un experimento como ejemplar sobre el análisis predictivo aplicado por regresión lineal en Microsoft Azure en la guía a presentar, para demostrar la precisión y fácil manejo del Studio ML.
- Presentar los niveles de consumo del Studio Machine Learning por usuario, por hora, y por transacciones de API's.

OBJETIVOS ESPECÍFICOS DE RESOLUCIÓN DEL PROBLEMA

- Diseñar una guía para los estudiantes, dando a conocer el modelo de Regresión Lineal en Machine Learning de Microsoft Azure, su funcionamiento y sus características.

JUSTIFICACIÓN

Debido a la gran cantidad de herramientas informáticas que salen cada día, una más completa que la otra y todas con el fin de simplificar tareas y hacer la vida más fácil, surge la necesidad de estar al día y contar con el conocimiento de las más usadas y por supuesto las que se adapten más a nuestras necesidades, con el propósito de alcanzar competencias y experiencias que se puedan aplicar como estudiante o profesional. Lo determino un requisito para alcanzar una situación laboral de éxito, el estar siempre actualizado.

El Cloud Computing es un modelo o paradigma para la utilización de los recursos informáticos, que está completamente basado en Internet. Dichos recursos, que normalmente son aplicaciones software, pero que también pueden ser almacenamiento de archivos, bases de datos, correo electrónico, etc., residen en servidores remotos, de modo que el usuario puede acceder a ellos desde cualquier lugar del mundo, siempre que cuente con un navegador y una conexión a Internet. (CIBERNAT, 2014)

Este paradigma está cambiando todo lo que conocíamos anteriormente, existen un montón de aplicaciones actualmente basadas en la nube y ha sido el avance que necesitaba la tecnología de la información, naciendo de éste Microsoft Azure, la plataforma de Servicios orientado a todo el que quiera crear, implementar, administrar aplicaciones y realizar pronósticos. El machine Learning es una de las tecnologías que ofrece esta plataforma, el poder de su conocimiento y de lo que brinda como sus modelos y características, debe ser aprovechado y enseñado, para abrir un mundo de oportunidades. Dar valor a nuestras aplicaciones web de forma más rápida, con algoritmos comprobados y métodos estadísticos-físicos avanzados más concisos.

Por tanto, este proyecto se justifica por los siguientes motivos:

- Permitir a los estudiantes el uso del modelo de Regresión Lineal en Machine Learning para alcanzar sus competencias de manera más ágil mediante Azure.
- Facilitar el uso correcto de la herramienta y su alcance por medio de un instructivo al alcance de todos.
- Asegurar un mejor aprendizaje e interés con nuevos métodos.
- Fomentar el uso de tecnología actual con un enfoque hacia el emprendimiento y hacia una vida laboral exitosa.

IMPACTOS ESPERADOS

IMPACTO TECNOLÓGICO

Debido al crecimiento de la tecnología junto a la inteligencia artificial, se han incrementado herramientas y aplicaciones cada vez más inteligentes que ayudan a llevar una vida sin complicaciones, de tal manera que ahora son usadas por las empresas y por un sin número de instituciones que desean obtener mayor rendimiento y mejores predicciones en cuanto a costo-tiempo-beneficio.

El uso de estas herramientas permite conocer el futuro y además seguirnos desarrollando en cuanto a aprendizaje que es lo que necesita la tecnología, mayor investigación que multiplique proyectos y nuevos inventos.

IMPACTO SOCIAL

Azure nace de la necesidad que se ha tenido al no contar con el dinero suficiente para tener equipos veloces con altas tecnologías o toda la infraestructura deseada en poco tiempo, y el Microsoft Azure Aprendizaje Automático va de la mano con esta filosofía.

Hacer un estudio como el que realiza Machine Learning o también conocido como Aprendizaje Automático antes costaba mucho y tomaba demasiado tiempo, pero ahora es posible por su disponibilidad y fácil acceso.

Los empresarios, los estudiantes pueden sacar partido totalmente de los servicios que encuentran y hacer implementaciones o construir software y más de manera fácil y rápida. El aprendizaje automático es un servicio que da valor a todo lo que toca, por no ser común, desde el momento que se habla de inteligencia artificial todos sabemos que es algo importante, y si se le saca el potencial se pueden hacer cosas sorprendentes con arduo aprendizaje y con la investigación que se merece.

CAPITULO I: MARCO TEÓRICO DE LA INVESTIGACIÓN

1.1 INTRODUCCIÓN

En este capítulo se encontrará el desarrollo y descripción de un conjunto de conceptos, definiciones e hipótesis seleccionadas, para hacer referencia a la solución planteada en este proyecto. Esta composición permitirá la comprensión sobre el Cloud Computing y Azure como proveedora de servicios en la nube: sus bases, el entorno en el que se desenvuelve, su funcionamiento, entre otras definiciones claves para poder usar esta plataforma, sacarle su potencial y hacerla una parte fundamental en el funcionamiento de nuestros proyectos académicos y profesionales futuros.

1.2. ANTECEDENTES DE INVESTIGACIONES RELACIONADAS AL TEMA.

INVESTIGACIONES ACERCA DE MICROSOFT AZURE Y CLOUD COMPUTING.

En la tesis sobre el “Análisis de la plataforma Windows Azure para trabajar en entornos cloud y su aplicación en la ESPOCH”, publicada en la página web de la ESPOCH, en el año 2011, René Omar Villa López y Liliana Grimaneza Guzmán Acán, idearon una investigación con conceptos y formas para implementar una aplicación web de información en Azure, en este caso para la Asociación de la Escuela de Ingeniería De Sistemas, en las que obtuvieron como resultado no sólo el software sino también, conclusiones como que el análisis del costo y el tiempo que sostuvo el desarrollo de ésta, demostró que la utilización de la plataforma Windows Azure es la mejor opción en cuanto a resultados y precio, que gracias a la gran documentación de ésta fue más fácil su uso, tomando en cuenta por supuesto que se debe tener conocimientos sobre lógica de programación, bases de datos y frameworks cuando de desarrollo de software

se trata; y que la tendencia del internet es cada día más grande, razón por la cual desarrollar aplicaciones destinadas a la nube es lo mejor.

En la tesis con el tema “Microsoft Windows Azure Como Plataforma Para Prestación de Servicios, Soluciones y Computación en la Nube”, alojada en la página web de la Universidad de Israel de la Ciudad de Quito, del autor Rikelmer Mamfredy Mejía Matute, presentada en el año 2012, se encontró un estudio sobre las características más usadas de Azure, así como de la instalación de la plataforma que en ese tiempo aún se hacía y cómo funcionaba por detrás. La finalidad de este proyecto era presentar una guía de gestión y administración de Microsoft Windows Azure en sus servicios más utilizados, se logró observar los cambios que ha tenido MA a lo largo del tiempo y que al igual que la tesis anterior, coinciden en que es una plataforma excelente tomando en cuenta la comparación costo-beneficio con otras propuestas de Cloud Computing.

En la tesis con el tema “Arquitecturas de red para servicios en Cloud computing”, alojada en la página web de la Universidad Complutense en Madrid, de los autores Jorge Lastras Hernansanz, Javier Lázaro Requejo y Jonatan David Mirón García, presentada en el año 2009, se pudo ver que querían implementar un servidor web a partir de la unión de recursos virtualizados de diferentes dominios basándose en el Cloud Computing, se dieron cuenta que se podían aprovechar recursos por medio de los servicios que ofrecían las diferentes plataformas de servicios como Azure, AWS, entre otros. La implementación puede requerir un día de que se dispongan de ciertos recursos, y de acuerdo al crecimiento puede ir aumentando, por lo que ellos notaron que pueden hacer escalables sus implementaciones con este tipo de plataformas al no desaprovechar infraestructura y también al satisfacer una mayor demanda del servicio si lo exige.

INVESTIGACIONES ACERCA DE MACHINE LEARNING.

En la tesis de masterado Appnalyze, sobre clasificación de comentarios de aplicaciones móviles, el 29 de Mayo del año 2015, Salvador Medina Herrera estudió la aplicación de técnicas de Machine Learning, en particular de diferentes modelos de Support Vector Machines, para clasificar conjuntos de comentarios de aplicaciones. Estos comentarios fueron extraídos de tiendas online de aplicaciones para móviles. En particular, utilizó técnicas de aprendizaje automático supervisado y comparó diferentes modelos capaces de decidir si cada uno de los comentarios analizados contiene o no oraciones con información subjetiva sobre la aplicación, valorando la aplicación en cuestión. Además, asignó cada una de estas valoraciones a cero o más temas generales, como la facilidad de uso, la estabilidad, la funcionalidad, el diseño y el rendimiento, así como la inferencia de la polaridad de dichas valoraciones, es decir, descubrir si se trata de comentarios positivos o negativos. Según los resultados obtenidos confirmaron que la aplicación de clasificadores basados en SVM resultan perfectamente viable para identificar tanto el tema como la polaridad de comentarios para diversos tipos de aplicaciones móviles; sin embargo comprobaron que la efectividad se limita por condicionantes como: ámbitos de las aplicaciones, el excesivo número de tokens especialmente considerable cuando los tokens utilizados son de baja frecuencia y la clasificación de conceptos más generales como el hecho de que un documento sea o no subjetivo resulta más compleja que para el caso de conceptos más concretos. (Herrera, 2015)

En el artículo publicado en la revista Alergia México, sobre Análisis discriminante para predecir el diagnóstico clínico de inmunodeficiencias primarias, año 2015, Chiharu Murata, Ana Belén Ramírez, Guadalupe Ramírez, Alonso Cruz, José Luis Morales y Saul Oswaldo Lugo-Reyes identificaron las

características que mejor explican la pertenencia de pacientes pediátricos con inmunodeficiencias primarias a un grupo de defectos o a una enfermedad por medio del reconocimiento de patrones. Las inmunodeficiencias primarias son un grupo heterogéneo de más de 250 enfermedades congénitas con mayor susceptibilidad a padecer infecciones, autoinflamación, autoinmunidad, alergia y cáncer. El análisis discriminante lineal es un método multivariante de clasificación supervisada para agrupar a los sujetos a partir de encontrar combinaciones lineales de un número de variables; partiendo de esto hicieron un estudio analítico transversal en el que a partir de una base de datos preexistente, con registros clínicos y de laboratorio de 168 pacientes con inmunodeficiencia primaria, seguidos en el Instituto Nacional de Pediatría de 1991 a 2012, construyeron discriminantes lineales para explicar la pertenencia de cada paciente a los diferentes grupos de defectos y a las inmunodeficiencias primarias más prevalentes en su registro.

Prácticamente en todos los casos el desempeño de la máquina fue superior al del experto humano en lo que respecta a la selección de los atributos más pertinentes para incorporar en los modelos. La predicción del diagnóstico con base en las ecuaciones construidas tuvo exactitud global de 83 a 94%, con sensibilidad de 60 a 100%, especificidad de 83 a 95% y coeficiente kappa de 0.37 a 0.76.

Concluyeron la investigación como un primer paso en la construcción de un sistema de aprendizaje automático, con un abordaje más amplio que incluya una base de datos más grande y diferentes metodologías, para asistir el diagnóstico clínico de las inmunodeficiencias primarias. (Chiharu Murata, 2015)

En el artículo publicado en el Blog dev academy, sobre Machine Learning e Inteligencia Artificial, el 12 de Marzo del año 2014, Pedro Muñoz del Río quién está haciendo su tesis doctoral sobre el tema, habló por el medio de difusión Hangout sobre los diferentes algoritmos de Machine Learning, sobre lenguajes

como Python, R y Weka, así como sus experiencias haciendo una tesis que trata sobre la aplicación de un modelo de regresión lineal para predecir el valor de empresas mineras a futuro en el sector financiero por medio de Weka. Se puede respecto a la conferencia sacar algunas conclusiones, una de ellas que el Machine Learning tiene muchas utilidades, y que lo más difícil es reunir la data y limpiar el ruido que esta tiene, debido a la propia experiencia de Pedro; quien se ingenió el método para la obtención de data que necesitaba para su demostración, primero escribiendo en Lenguaje R, unos scripts que obtuvieran datos de la bolsa de valores, y luego delimitando su tema, con empresas mineras porque tenía mayor acceso a data. Otro punto también definido es la necesidad de conocer bien la data con la que se va a trabajar, para obtener mejores resultados al momento de generar una predicción.

En el artículo publicado en el Blog Koalite, sobre Machine Learning con Princesas Disney, del autor Juan María Hernández, año 2015, se encuentra una interesante perspectiva de lo que es Machine Learning y como su mayor evolución ha sido ahora, debido a la nube y el volumen de datos recolectado por los sistemas modernos. Comenta sobre la potencia de cálculo con más accesibilidad y el poder de contar con grandes volúmenes de datos sofisticados para la alimentación de algoritmos.

El ML está empezando a usarse cada vez más, y aunque su base matemática subyacente es un poco compleja, con librerías como numl se la puede usar de manera más fácil.

Demuestra a través de esta librería, que aunque parezca que el modelo que usemos para hacer machine Learning se construye por algoritmos sofisticados muy inteligentes y que eso va a significarlo todo, en realidad eso no va a determinar la exactitud de nuestra predicción, nuestro análisis inicial es el que va a condicionar realmente la utilidad de los resultados a obtener.

En el artículo publicado en la página web CleverTask, con el tema “De los datos a las predicciones”, del autor Andrés González, año 2015, se encuentra una aportación importante, un ejemplo en su máxima expresión lógica, que se puede aplicar a la vida real, en este caso en el sector turístico de España, en el cual explican las etapas para hacer un Machine Learning para predecir esta situación, es un ejemplo técnico y a la vez simple, que lo muestra como una serie de pasos a seguir o su parte lógica si se la quiere ver desde otro perfil. Se empieza con la captura de datos en la que se consigue toda la información pasada que esté directa e indirectamente conectada con los datos a predecir, la limpieza de los valores que produzcan ruido como son los valores con null, las transformaciones que se dan en la data como lo es la transformación a formatos como csv, la selección de los datos de calidad como son las columnas de una tabla que sean relevantes para la predicción, la creación del modelo de predicción como pueden ser los modelos de tasa de acierto, matriz de confusión, entre otros, para saber el entrenamiento que requiere y el porcentaje que define la calidad del modelo.

1.3 DEFINICIONES CONCEPTUALES

1.3.1 Aprendizaje.

Es el proceso de asimilación de información mediante el cual se adquieren nuevos conocimientos, técnicas o habilidades.

En este sentido, el aprendizaje consiste en adquirir, procesar, entender y aplicar una información que nos ha sido enseñada o que hemos adquirido mediante la experiencia a situaciones reales de nuestra vida. De allí que el aprendizaje pueda observarse tanto en los seres humanos como en los animales. (7G, 2013-2015)

El aprendizaje es el resultado de todos los procesos de estudio, análisis y experiencias que el ser vivo va obteniendo a lo largo de su vida, lo que altera el comportamiento que actúa de acuerdo al conocimiento.

1.3.1.1 Inteligencia.

La inteligencia es la capacidad de asimilar, guardar, elaborar información y utilizarla para resolver problemas, cosas que también son capaces de hacer los animales e incluso los ordenadores. Pero el ser humano va más allá, desarrollando una capacidad de iniciar, dirigir y controlar nuestras operaciones mentales y todas las actividades que manejan información. Aprendemos, reconocemos, relacionamos, mantenemos el equilibrio y muchas cosas más sin saber cómo lo hacemos. Pero tenemos además la capacidad de integrar estas actividades mentales y de hacerlas voluntarias, en definitiva de controlarlas, como ocurre con nuestra atención o con el aprendizaje, que deja de ser automático como en los animales para focalizarlo hacia determinados objetivos deseado. (PsicoActiva, 1998-2015)

1.3.1.2 Inteligencia artificial

La inteligencia artificial, por lo tanto, consiste en el diseño de procesos que, al ejecutarse sobre una arquitectura física, producen resultados que maximizan una cierta medida de rendimiento. Estos procesos se basan en secuencias de entradas que son percibidas y almacenadas por la mencionada arquitectura.

Los dispositivos que cuentan con inteligencia artificial pueden ejecutar distintos procesos análogos al comportamiento humano, como la devolución de una respuesta por cada entrada (similar a los actos reflejos de los seres vivos),

la búsqueda de un estado entre todos los posibles según una acción o la resolución de problema mediante una lógica formal.

(Definicion.de, 2008-2015)

1.3.2 Aprendizaje de Máquina - Machine Learning.

Es un conjunto de técnicas pertenecientes a la inteligencia artificial cuyo objetivo es permitir que las máquinas “aprendan”. De forma más específica se trata de construir y estudiar algoritmos que, en base al análisis de (grandes) conjuntos de información, permitan realizar predicciones o tomar decisiones. Esto implica que un software que utilice estas técnicas, no tiene por qué seguir un comportamiento predictivo en base a una secuencia de instrucciones escritas por el programador, sino que tendrá comportamientos y resultados variables en base a algoritmos que analizan la información y construyen un modelo de acciones en tiempo real. (Pallo, 2014)

1.3.2.1 Conceptos Machine Learning

- **Minería de datos (data mining):** La minería de datos y machine learning. Son conceptos primos hermanos. La principal diferencia radica en el objetivo que tiene cada una de las disciplinas. Mientras que la minería de datos descubre patrones anteriormente desconocidos, el aprendizaje automático se usa para reproducir patrones conocidos y hacer predicciones basadas en los patrones.
- **Aprendizaje o entrenamiento (learning, training):** Proceso que detecta los patrones de un conjunto de datos, el corazón del machine learning. Una vez identificados los patrones, se pueden hacer predicciones con nuevos datos que se incorporen al sistema.
- **Conjunto de datos (dataset):** Es la materia prima del sistema de predicción. Es el histórico de datos que se usa para entrenar al sistema

que detecta los patrones. El conjunto de datos se compone de instancias, y las instancias de factores, características o propiedades.

- **Instancia, ejemplo o registro (instance, sample, record):** Una instancia es cada uno de los datos de los que se disponen para hacer un análisis. Si se quiere predecir el comportamiento de los clientes de un servicio de telefonía, cada instancia correspondería a un abonado. Cada instancia, a su vez, está compuesta de características que la describen.
- **Característica, atributo, factor, propiedad o campo (feature, attribute, property, field):** Atributos que describen cada una de las instancias del conjunto de datos.
- **Objetivo (objective):** Es el atributo o factor que queremos predecir, el objetivo de la predicción.
- **Ingeniería de factores (feature engineering):** Proceso previo a la creación del modelo de predicción en el que se hace un análisis, limpieza y estructuración de los campos de los datos. Este proceso es uno de los más importantes y más costosos del proceso de predicción. Proceso que elimina el ruido de la señal.
- **Modelo (model):** Tras entrenar al sistema, se crea un modelo que servirá para hacer las predicciones.
- **Árbol de decisión (decision tree):** Representación del modelo en modo gráfico como un árbol, en el que las ramas constituyen los patrones reconocidos en el proceso de aprendizaje. En las hojas de las ramas se situarían las predicciones para cada patrón.
- **Confianza (confidence):** Es la probabilidad de acierto que calcula el sistema para cada una de las predicciones.

(Gonzalez, 2014)

1.3.2.2 Áreas de Machine Learning

a) Aprendizaje no supervisado.

Cuando no hay un conocimiento a priori de las salidas que corresponden al conjunto de datos de entrada. En estos casos el objetivo es encontrar grupos mediante **clustering** o determinar una **distribución de probabilidad** sobre un conjunto de entrada.

- Clustering: Dada una colección de datos de clientes, encontrar una forma de agruparlos, dado distintas variables, como ingreso, genero, compras, etc.
- Asociación: Un doctor, con muchos años de experiencia, puede encontrar una asociación entre un conjunto de características, síntomas, registro históricos y una enfermedad. (Brange, 2013)

b) Aprendizaje supervisado.

Cuando el conjunto de datos viene con los atributos adicionales que queremos predecir. El problema puede clasificarse en dos categorías:

- Regresión: los valores de salida consisten en una o más variables continuas. Un ejemplo es la predicción del valor de una casa en función de su superficie útil, número de habitaciones, cuartos de baños, etc.
- Clasificación: las muestras pertenecen a dos o más clases y queremos aprender a partir de lo que ya conocemos cómo clasificar nuevas muestras. (Brange, 2013)

1.3.3 Tecnología.

Siguiendo su concepto etimológico puede considerarse su definición: el arte o técnica que, a través conocimientos científicos aplicados de forma lógica y

ordenada, permite la fabricación de objetos, artefactos, herramientas o instrumentos que sirven para cubrir las necesidades de la persona o grupos de personas. Es un proceso combinado de pensamiento y acción con la finalidad de generar soluciones que sean útiles. (Quees.la, 2015)

El significado de tecnología podría interpretarse como el estudio de técnicas, lo que deja paso a la siguiente definición:

“La tecnología es un concepto amplio que abarca un conjunto de técnicas, conocimientos y procesos, que sirven para el diseño y construcción de objetos para satisfacer necesidades humanas” (Alegsa, ALEGSA.com.ar, 2015)

Podemos decir entonces, que la tecnología es la creación de herramientas que ayudan y facilitan la realización de una tarea específica, satisfaciendo así, una necesidad humana; sin embargo es necesario añadir que sólo será denominada así, mientras su técnica se vincule con la ciencia, y al hacerla se le sume la teoría.

1.3.4 Tecnología de la información.

Se conoce como tecnología de información (TI) a la utilización de tecnología específicamente computadoras y ordenadores electrónicos para el manejo y procesamiento de información como la captura, transformación, almacenamiento, protección, y recuperación de datos e información. (Meltom Technologies, n.d.)

Actualmente la podemos encontrar en cualquier lugar, desde que abrimos nuestra bandeja de mensajes, publicamos algo en Facebook, guardamos archivos en nuestro OneDrive o hasta cuando sacamos dinero por el cajero.

1.3.4.1 Almacenamiento de Datos

Para cualquier sistema ordenado, las unidades de almacenamiento son aquellas que permiten guardar física o virtualmente archivos de datos de todo tipo. (definicion abc, 2007-2015)

El almacenamiento de datos es fundamental para la tecnología como tal, sin éste, el mundo no existiera como lo conocemos porque gracias a los dispositivos de almacenamiento podemos acceder, procesar o manipular información. Es igual de importante para las empresas, porque ésta es la base de su negocio, y de la buena o mala administración que se le dé a los datos, dependerá el éxito que tenga.

a) Bases de Datos

Es el conjunto de informaciones almacenadas en un soporte legible por ordenador y organizadas internamente por registros (formado por todos los campos referidos a una entidad u objeto almacenado) y campos (cada uno de los elementos que componen un registro). Permite recuperar cualquier clase de información: referencias, documentos textuales, imágenes, datos estadísticos, etc. (Molina, 2011)

Una base de datos es una colección de información organizada de forma que un programa de ordenador pueda seleccionar rápidamente los fragmentos de datos que necesite. Una base de datos es un sistema de archivos electrónico. (Masadelante.com, 2015)

- **Bases de Datos relacionales:** Una base de datos relacional es una base de datos en donde todos los datos visibles al usuario están organizados estrictamente como tablas de valores, y en donde todas las operaciones de la base de datos operan sobre estas tablas. Estas bases de datos son percibidas por los usuarios como una colección de relaciones normalizadas de diversos grados que varían con el tiempo. El modelo

relacional representa un sistema de bases de datos en un nivel de abstracción un tanto alejado de los detalles de la máquina subyacente. (Miranda, 2007)

- **Bases de Datos noSQL:** acrónimo de Not Only SQL, que aboga más por la evolución del Internet; no es exactamente un tipo de base de datos, sino un conjunto de tipos de bases de datos, por ejemplo, con las bases de datos documentales que son las más conocidas, se podría hacer prácticamente todo lo que se hace actualmente con una relacional. Cabe destacar que en NoSQL no existe ACID (Atomicity, Consistency, Isolation, Durability) como en las bases de datos relacionales, pero existe algo llamado BASE (Basic Availability, Soft State, Eventual Consistency) que da características para el manejo de datos. (Graterol, 2014)

b) **Redundancia de Datos**

La tolerancia de fallos se basa en la redundancia. Todas las técnicas de tolerancia a fallos confían en elementos adicionales introducidos en el sistema para detectar el fallo y recuperarse del mismo. El objetivo de la tolerancia a fallos es minimizar la redundancia y maximizar la fiabilidad proporcionada sujeta a las restricciones de costo y tamaño del sistema y puede introducir fallos adicionales. Por esta razón es conveniente aislar los componentes redundantes del resto del sistema.

Se utilizan técnicas distintas de redundancia para hardware y software.

Para el hardware:

- **Redundancia estática:** Los componentes redundantes están siempre activos. Se utilizan para ocultar o enmascarar los fallos. Una de estas técnicas es la redundancia modular triple (RMT) o redundancia modular

enésima (RMN). Consiste en tres (n) subsistemas idénticos que operan en paralelo y un circuito votador. El resultado final se elige por mayoría. Con $n=3$ se enmascaran fallos en un componente, con $n=2$ se pueden detectar fallos, no enmascararlos.

- **Redundancia dinámica:** Los componentes redundantes se activan cuando se detecta un fallo. Ejemplos de redundancia dinámica son los “checksums” en los paquetes de comunicación y los bits de paridad en los accesos de memoria.

Para el software:

- **Redundancia estática:** el éxito del hardware RMT y RMN ha motivado una aproximación similar al software de tolerancia a fallos. Ahora bien, dado que el software no se deteriora como el hardware, la aproximación se utiliza para detectar fallos en el diseño. Consiste en desarrollar por equipos diferentes sin interacción n versiones independientes (3 como mínimo) del sistema basadas en la misma especificación. Las n versiones se activan concurrentemente y cuando terminan, se comparan sus resultados mediante un proceso guía. Se toma como resultado del conjunto uno sobre el que haya consenso, por ejemplo votación.
- **Redundancia dinámica:** Sólo se ejecutan cuando se detecta un error, se intenta dejar el sistema en un estado correcto y continuar la ejecución, aunque sea con menor funcionalidad. Esta técnica tiene 4 etapas: Detección de errores, Evaluación y confinamiento de los daños, Recuperación de errores y Reparación de fallos. (López, 2005)

1.3.4.2 Manipulación de Datos

Hay muchas formas de manipular la información. La manipulación de la información en Internet está relacionada con uno de los siguientes puntos: cómo se forma la información; usos "violentos" de las técnicas por Internet para manipular a los usuarios (a través de mensajes por correo electrónico o la publicidad "violenta") y los monopolios en Internet. (SCAS, 2010-2011)

Lo anteriormente dicho simplemente es algo que se puede lograr con la manipulación de información, pero ésta no tiene que afectar la seguridad de las personas o las empresas, es cierto que es preciso dar un buen uso; sin embargo siempre que sea buena tiene muchos beneficios.

Nació debido a que una vez almacenada mucha información, era poco visitada, problema que con el tiempo se solucionó con la manipulación de datos, que dio pie a la minería de Datos, así como al grandioso Big Data. La manipulación de datos es necesaria para poder consultar información de una manera más óptima, incluso para hacer proyecciones.

a) Recuperación de Datos

Técnicas empleadas para recuperar archivos que han sido perdidos o eliminados de algún medio de almacenamiento. (Alegsa, ALEGSA.com.ar, 2015)

Una pérdida de datos puede darse de manera física o lógica. Cuando se trata de una pérdida física es cuando se torna más difícil el recuperar la información, debido al problema que implica el que la materia física sobre la que está nuestra información ha sido perjudicada; pero si hablamos de una pérdida lógica si se la intenta recuperar con rapidez es muy probable que tenga solución.

Esto se debe a que cuando eliminamos un archivo de cualquiera de nuestros dispositivos de almacenamiento, este lo marca como eliminado para

que pueda ser sobrescrito con el tiempo, más no lo elimina en su totalidad como pensamos.

1.3.4.3 Transmisión de Datos.

Es un movimiento de transmisión donde se encuentra la información codificada de un punto o más puntos mediante señales eléctricas, ópticas, electroópticas o electromagnéticas. Este rendimiento originado en las organizaciones gubernamentales, industriales, comerciales, bancarias, empresariales, militares han nacido por la necesidad de poner a disposición de ellas en un punto remoto la capacidad de proceso de un ordenador, ubicado en un punto se podría llamar central. (esthelar, 2008)

Cuando se desea compartir información se necesitan los medios de transmisión para que esta viaje. Se hará por medio guiados o no guiados, es decir con cable o sin cable respectivamente. Su objetivo es transmitir los datos de manera rápida y con calidad, con disminución de tiempo y esfuerzo.

1.3.4.4 Centro de Datos (Datacenter)

Un data center también llamado centro de datos es un espacio acondicionado especialmente para contener a todos los equipos y sistemas de TI.

Cuando indico especialmente acondicionado me refiero a que es un lugar que tiene instalado lo siguiente: climatización (aire acondicionado), alimentación eléctrica estabilizada e ininterrumpida, cableado estructurado, sistemas contra incendios, control de acceso, sistemas de cámaras de vigilancia, alarmas contra incendios, control de temperatura y humedad. (aprendaRedes.com, 2015)

1.3.5 Arquitectura cliente/Servidor.

La tecnología Cliente/Servidor es el procesamiento cooperativo de la información por medio de un conjunto de procesadores, en el cual múltiples clientes, distribuidos geográficamente, solicitan requerimientos a uno o más servidores centrales.

Desde el punto de vista funcional, se puede definir la computación Cliente/Servidor como una arquitectura distribuida que permite a los usuarios finales obtener acceso a la información de forma transparente aún en entornos

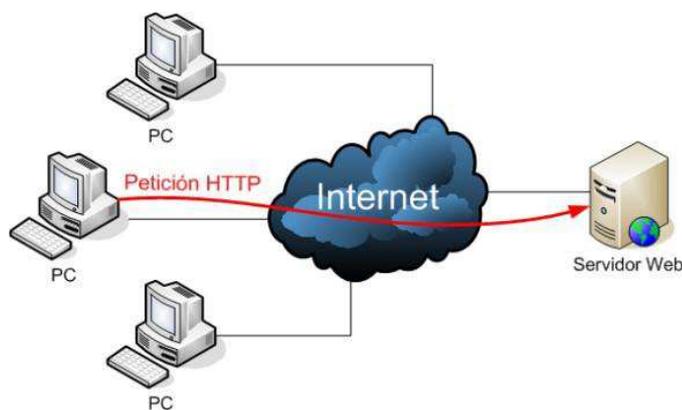


Ilustración 1 Arquitectura cliente-servidor

multiplataforma. Se trata pues, de la arquitectura más extendida en la realización de Sistemas Distribuidos. (Ángel, Mialto, 2015)

1.3.5.1 Características

Un sistema Cliente/Servidor es un Sistema de Información distribuido basado en las siguientes características:

- Servicio: unidad básica de diseño. El servidor los proporciona y el cliente los utiliza.

- Recursos compartidos: Muchos clientes utilizan los mismos servidores y, a través de ellos, comparten tanto recursos lógicos como físicos.
- Protocolos asimétricos: Los clientes inician “conversaciones”. Los servidores esperan su establecimiento pasivamente.
- Transparencia de localización física de los servidores y clientes: El cliente no tiene por qué saber dónde se encuentra situado el recurso que desea utilizar.
- Independencia de la plataforma HW y SW que se emplee.
- Sistemas débilmente acoplados. Interacción basada en envío de mensajes.
- Encapsulamiento de servicios. Los detalles de la implementación de un servicio son transparentes al cliente.
- Escalabilidad horizontal (añadir clientes) y vertical (ampliar potencia de los servidores).
- Integridad: Datos y programas centralizados en servidores facilitan su integridad y mantenimiento.
- En el modelo usual Cliente/Servidor, un servidor, (daemon en la terminología sajona basada en sistemas UNIX/LINUX, traducido como “demonio”) se activa y espera las solicitudes de los clientes. Habitualmente, programas cliente múltiples comparten los servicios de un programa servidor común. Tanto los programas cliente como los servidores son con frecuencia parte de un programa o aplicación mayores. (Ángel, Mialto, 2015)

1.3.5.2 Modelo Cliente/Servidor

El modelo Cliente/Servidor es un modelo basado en la idea del servicio, en el que el cliente es un proceso consumidor de servicios y el servidor es un proceso proveedor de servicios. Además esta relación está establecida en

función del intercambio de mensajes que es el único elemento de acoplamiento entre ambos.

De estas líneas se deducen los tres elementos fundamentales sobre los cuales se desarrollan e implantan los sistemas Cliente/Servidor: el proceso cliente que es quien inicia el diálogo, el proceso servidor que pasivamente espera a que lleguen peticiones de servicio y el middleware que corresponde a la interfaz que provee la conectividad entre el cliente y el servidor para poder intercambiar mensajes. (Ángel, Mialto, 2015)

a. ¿Qué es un cliente?

Un cliente es todo proceso que reclama servicios de otro. Una definición un poco más elaborada podría ser la siguiente: cliente es el proceso que permite al usuario formular los requerimientos y pasarlos al servidor. Se lo conoce con el término front-end.

Las funciones que lleva a cabo el proceso cliente se resumen en los siguientes puntos:

- Administrar la interfaz de usuario.
- Interactuar con el usuario.
- Procesar la lógica de la aplicación y hacer validaciones locales.
- Generar requerimientos de bases de datos.
- Recibir resultados del servidor.
- Formatear resultados.

(Ángel, Mialto, 2015)

b. ¿Qué es un middleware?

El middleware es un módulo intermedio que actúa como conductor entre sistemas permitiendo a cualquier usuario de sistemas de información comunicarse con varias fuentes de información que se encuentran conectadas por una red. En el caso que nos concierne, es el intermediario entre el cliente y el servidor y se ejecuta en ambas partes. (Ángel, Mialto, 2015)

c. ¿Qué es un servidor?

Un servidor es todo proceso que proporciona un servicio a otros. Es el proceso encargado de atender a múltiples clientes que hacen peticiones de algún recurso administrado por él. Al proceso servidor se lo conoce con el término back-end. El servidor normalmente maneja todas las funciones relacionadas con la mayoría de las reglas del negocio y los recursos de datos. Puede darse el caso que un servidor actúe a su vez como cliente de otro servidor.

Las principales funciones que lleva a cabo el proceso servidor se enumeran a continuación:

- Aceptar los requerimientos de bases de datos que hacen los clientes.
- Procesar requerimientos de bases de datos.
- Formatear datos para transmitirlos a los clientes.
- Procesar la lógica de la aplicación y realizar validaciones a nivel de bases de datos.

(Ángel, Mialto, 2015)

"Server" ó servidor, también llamado "Host" ó anfitrión; es una computadora especializada con muy altas capacidades de proceso, encargada de proveer diferentes servicios a las redes de datos (una red es un conjunto de computadoras interconectadas entre sí), tanto inalámbricas como las basadas en

cable; también permite accesos a cuentas de correo electrónico, administración de dominios empresariales, hospedaje y dominios Web entre otras funciones. (InformaticaModerna.com, 2008-2015)

Se debe tener en cuenta que aunque un servidor puede entregar algunos servicios, no es lo más aconsejable por cuestiones de seguridad, porque si se diera un problema de invasión o ataque al servidor sólo sería afectado un servicio y no todos, aminorando caídas de sistemas a grandes escalas. Casi cualquier computador puede funcionar como servidor; sin embargo al contar con servidores dedicados siempre serán más adecuados para un ambiente de producción grande.

Los tipos de servidor más comunes son los de: archivos, base de datos, correo, impresión, web, juegos y aplicaciones.

- **Servidor de Archivos**

El servidor de archivos permite que cualquier usuario de nuestra red suba allí un determinado archivo, que será accesible desde Internet, y que podrá ser descargado por cualquiera que conozca la dirección desde donde descargarlo (a esto se le llama enlace, link, URL, dirección, etc.). El archivo se sube por única vez, y se transfiere solo a los destinatarios que están interesados en obtener una copia. El servidor guarda registro de la cantidad de veces que se descargó cada archivo, y la última fecha de descarga del mismo. (Pasqualini, 2010)

Es uno de los servidores más usados en la actualidad por ser un elemento clave en la administración y el flujo de documentos en las empresas.

- **Servidor de Base de Datos**

Un servidor de bases de datos se utiliza para almacenar, recuperar y administrar los datos de una base de datos. El servidor gestiona las

actualizaciones de datos, permite el acceso simultáneo de muchos servidores o usuarios web y garantiza la seguridad y la integridad de los datos. Y cuando hablamos de datos, podemos estar hablando sobre millones de elementos a los que acceden al mismo tiempo miles de usuarios. (Turner, 2014)

Los servidores de bases de datos no sólo permiten la actualización de nuestros datos, sacar respaldos o asegurar su integridad, también cumplen un papel relevante al momento de ser uno de los elementos principales, usados para el alojamiento de nuestras aplicaciones web. Es el abastecedor de requerimientos solicitados por los usuarios por medio de las diferentes interfaces, en cuanto a manejo de información se trata. Si un servidor de base de datos está caído, nuestra aplicación no tardará en colapsar también, al momento en que se haga una llamada a éste.

- **Servidor de Correo**

Un servidor de correo es una aplicación informática que tiene como objetivo, enviar, recibir y gestionar mensajes a través de las redes de transmisión de datos existentes, con el fin de que los usuarios puedan mantenerse comunicados con una velocidad muy superior a la que ofrecen otros medios de envío de documentos. (Culturación, n.d.)

Es uno de los servicios más utilizados por usuarios en computadores mundialmente, la mayoría de habitantes del planeta mayor a 18 años tiene por lo menos una cuenta de correo electrónico, por ello en toda organización es importante contar con un servidor de correo, que asegure el envío íntegro de información y velocidad de transmisión excelente. Existen un sin número de servidores de correos, entre los más conocidos están Exchange, Sendmail y Lotus; si hablamos de servidores de correos web, aunque en realidad son más bien clientes de correo que corren en un servidor web encontramos: www.hotmail.com, www.gmail.com, entre otros.

- **Servidor de Impresión**

Un servidor de impresión, o Print Server como también se lo conoce, es un pequeño dispositivo que podemos conectar a cualquier puerto disponible en el router o modem, y de este modo hacer accesible cualquier impresora que conectemos a él desde todas las impresoras que sean parte de la red, es decir que básicamente permitirá a las computadoras en una red acceder a una misma impresora. (INFORMÁTICA HOY, 2008)

Es un pequeño dispositivo útil tanto a nivel de hogar como empresarial, que al conectarse al router, al wi-fi, entre otras permite la conexión remota, lo que nos representa tiempo adicional al ahorrarnos la movilización y la instalación de software en nuestro computador, así como evitarnos la necesidad de prender la computadora a la que teníamos conectada la impresora como pasa con las impresoras compartidas.

- **Servidor Web**

Un servidor web es un programa que se ejecuta continuamente en un computador, manteniéndose a la espera de peticiones de ejecución que le hará un cliente o un usuario de Internet. El servidor web se encarga de contestar a estas peticiones de forma adecuada, entregando como resultado una página web o información de todo tipo de acuerdo a los comandos solicitados.

En este punto es necesario aclarar lo siguiente: mientras que comúnmente se utiliza la palabra servidor para referirnos a una computadora con un software servidor instalado, en estricto rigor un servidor es el software que permite la realización de las funciones descritas. (MISRESPUESTAS.COM, 2005-2015)

- **Servidor de juegos**

Servidor de juego online que emplea la conexión de un jugador, denominado anfitrión, para conectar entre sí a todos los jugadores participantes.

Es el tipo de servidor más empleado debido a su bajo coste, aunque tiene la desventaja de depender totalmente de la velocidad y estabilidad de la conexión del jugador anfitrión para gestionar la partida. (GamerDic, 2013-2015)

En el mundo del gaming es habitual que una comunidad se plantee la idea de alquilar uno o más servidores dedicados para jugar a juegos multijugador. Esto se debe a las características de determinados videojuegos, ya que para que se pueda jugar bien en línea y con muchos usuarios a la vez, hacen falta una serie de recursos que no te ofrece un hosting normal ni de lejos. Por eso se utilizan servidores dedicados en la inmensa mayoría de los casos. (Rico, 2015)

- **Servidor de aplicaciones**

Tipo de servidor que permite el procesamiento de datos de una aplicación de cliente. Las principales ventajas de la tecnología de los servidores de aplicación es la centralización y la disminución de la complejidad del desarrollo de aplicaciones, dado que las aplicaciones no necesitan ser programadas; en su lugar, estas son ensambladas desde bloques provistos por el servidor de aplicación. (WeBlog, n.d.)

Las principales ventajas de la tecnología de los servidores de aplicación es la centralización y la disminución de la complejidad del desarrollo de aplicaciones, dado que las aplicaciones no necesitan ser programadas; en su lugar, estas son ensambladas desde bloques provistos por el servidor de aplicación. (ALEGSA.com.ar, 2015)

1.3.6 Aplicaciones Informáticas.

Para la informática, una aplicación es uno de diversos tipos de programas de computación diseñados especialmente para complementar una función o actuar como herramienta para acciones puntuales del usuario.

A diferencia de otros programas como los sistemas operativos, los lenguajes de programación y otros, la aplicación tiene el único y principal fin de realizar una tarea específica, a menudo básica y de rápido y fácil uso para el usuario común, no avanzado. (definicionabc, 2015)

1.3.6.1 Aplicaciones de Escritorio.

Es la aplicación creada para ejecutarse en un ordenador de escritorio, sobre un sistema operativo de interfaz visual como Windows o Linux. (Lara, n.d.)

Antes del nacimiento de la web 2.0 era lo único que había. Ahora existen opciones que nos permiten tener centralizada la información y mayor integración; sin embargo depende de la necesidad usarlas. Una aplicación de escritorio te ofrece más confiabilidad, porque es más probable que la conexión a internet se encuentre interrumpida a que no tengamos acceso a nuestro computador, y en cuestiones de seguridad es la mejor, sin embargo al no tener centralizada la información se podría perder. Entonces es cuestión de perspectiva.

1.3.6.2 Aplicaciones Web.

Las aplicaciones web reciben este nombre porque se ejecutan en la internet. Es decir que los datos o los archivos en los que trabajas son procesados y almacenados dentro de la web. Estas aplicaciones, por lo general, no necesitan ser instaladas en tu computador.

El concepto de aplicaciones web está relacionado con el almacenamiento en la nube. Toda la información se guarda de forma permanente en grandes servidores de internet y nos envían a nuestros dispositivos o equipos los datos que requerimos en ese momento, quedando una copia temporal dentro de nuestro equipo. (GCF Community Foundation International, 1998-2015)

Son muy usadas actualmente debido a sus interfaces fáciles de adaptar, información centralizada y más que todo a la facilidad de actualizarlas y mantenerlas. Sin mencionar la versatilidad con la que cuentan al correr en cualquier sistema operativo.

1.3.6.3 Aplicaciones Móviles.

Una App es una aplicación de software que se instala en dispositivos móviles o tablets para ayudar al usuario en una labor concreta, ya sea de carácter profesional o de ocio y entretenimiento.

El objetivo de una app es facilitarnos la consecución de una tarea determinada o asistirnos en operaciones y gestiones del día a día. Existen infinidad de tipos de aplicaciones: Apps de noticias (elmundo.es), juegos (CSR Racing), herramientas de comunicación como Whatsapp, redes sociales como Google+, apps para salir de fiesta (Liveclubs), promociones comerciales (McDonalds)...etc, que nos pueden ayudar en el trabajo o intentar hacernos el día más ameno. (QODE, 2012)

Las aplicaciones móviles son fáciles de descargar gracias a las tiendas de las diferentes distribuciones, y por lo general gratuitas. Acceder a ellas es muy fácil, requiere sólo un celular que soporte y una conexión a internet, sea desde wi-fi o conexión de datos, para directamente realizar su descarga.

1.3.6.4 Software Utilitario.

Este tipo de programa o herramienta facilita las tareas del usuario en gran medida, debido a que no son grandes programas con un sin número de funcionalidades. Los utilitarios buscan la resolución de un problema en especial, para hacer de su funcionamiento algo sencillo.

Un utilitario es aquel software que está diseñado para realizar una tarea determinada, por ejemplo, un editor, un depurador de código o un programa recuperador de datos perdidos o borrados accidentalmente.

Se refiere al software que nos ayuda a resolver problemas relacionados con la administración de una computadora. (Rodriguez M. , 2012)

1.3.6.5 Sistemas Operativos.

El sistema operativo es el software, es decir un programa, definido como el más importante utilizados por las computadoras. Este software coordina y dirige servicios y aplicaciones que utiliza el usuario. Esto quiere decir que este programa permite que el resto de los programas funcionen ya que permite que se reconozcan ciertas conexiones, la seguridad de la computadora, realizar controles, envíos, etc. Los más utilizados son Windows, Linux, OS/2 y DOS. (CONCEPTO.DE, 2012-2015)

Podemos decir entonces que el sistema operativo, es prácticamente la configuración que necesita el computador para funcionar. Sin éste, ningún otro software funcionaría porque funciona como el intermediario entre el hardware (fierro) que vemos y las aplicaciones, todo esto gracias a sus tres componentes importantes: los sistemas de archivos, la interpretación de comandos y el núcleo.

1.3.7 Cloud Computing – Computación en la nube.

La computación en la nube, conocida también como informática en la nube, del inglés Cloud computing, es un modelo o paradigma para la utilización de los recursos informáticos, que está completamente basado en Internet. Dichos recursos, que normalmente son aplicaciones software, pero que también pueden ser almacenamiento de archivos, bases de datos, correo electrónico, etc., residen en servidores remotos, de modo que el usuario puede acceder a ellos desde cualquier lugar del mundo, siempre que cuente con un navegador y una conexión a Internet. (CIBERNAT, 2014)

1.3.7.1 Características.

1. Se vende por demanda, por lo general no por minutos ni horas, sino que es elástico, el usuario puede utilizar mucho o poco del servicio que desee en un momento dado. (ProspecNet S.L. , 2010)
2. El servicio está totalmente gestionado por el proveedor (el usuario necesita nada más que un dispositivo que sea capaz de navegar por internet). (ProspecNet S.L. , 2010)
3. Las innovaciones significativas en la virtualización y la computación distribuida, así como un mejor acceso a Internet de alta velocidad y la necesidad de reducir costos, son en parte los responsables del crecimiento acelerado del cloud computing. (ProspecNet S.L. , 2010)
4. El internauta puede en todo momento decidir qué aplicaciones usar y elegir entre aquellas que son gratuitas y las que no lo son. En el caso de las aplicaciones de pago, el coste irá en función de diversas variables, como el servicio contratado, el tiempo que se ha usado ese servicio, el volumen de tráfico de datos utilizado, etc. (Alcocer, 2010)
5. Las unidades de capacidad pueden ser rápidas y fácilmente aprovisionadas (en algunos casos en forma automática), escaladas

(crecimiento) o liberadas. Para el consumidor, estos recursos suelen parecer ilimitados, y pueden ser adquiridos en cualquier cantidad y momento. (Mesta, 2011)

1.3.7.2 Ventajas

- Menores Gastos Informáticos y de Software.
- Almacenamiento casi ilimitado.
- Reducción de impactos negativos en la ecología.
- Alta disponibilidad.
- Facilidades para trabajo en equipo.
- Actualizaciones del software.

1.3.7.3 Desventajas

- Requiere de conexión a internet permanente y de alta velocidad.
- Riesgos en la seguridad de la información.
- Riesgos de pérdida de privacidad.
- Dependencia del proveedor.

1.3.7.4 Tipos de Nube relevantes.

a) Nube Privada

Una nube privada es una nube en la que se aplica tecnología "cloud computing" pero con un modelo privado, es éste un medio en el que sólo una organización tiene acceso a los recursos que se utilizan para implementar la nube. En otras palabras, es una nube que una organización lleva a cabo con sus propios recursos (máquinas, redes, almacenamiento, centros de datos, etc) (K35 IT Managers Group Team, 2010)

b) Nube Pública

Una nube pública implementa el cloud computing para múltiples organizaciones, y estas, utilizan un único conjunto de recursos. El modelo básico para una nube pública es similar al de una compañía eléctrica pública: un tercero, administra la infraestructura necesaria para ofrecer capacidad de computación a los clientes que pagan por uso. (K35 IT Managers Group Team, 2010)

c) Nube Híbrida

Una nube híbrida es un tipo de modelo de implementación para el cómputo en la nube donde algunos recursos de TI se proporcionan como servicio a través de una nube privada interna y otros recursos son provistos por proveedores de servicios de terceros en la nube pública. Estos recursos de TI (aplicaciones, cómputo, almacenamiento y redes) se proporcionan según demanda a través de un catálogo de autoservicio con pago por el uso. (EMC, 2015)

1.3.7.5 Modelos de Servicio.

a) Plataforma como Servicio (PAAS)

Una plataforma como servicio proporciona una plataforma y un entorno para que el desarrollador cree sus propias aplicaciones y cuente con todos los servicios para que funcionen a través de internet.

El modelo PaaS permite a los usuarios crear aplicaciones de software utilizando herramientas suministradas por el proveedor. Los servicios PaaS pueden consistir en funcionalidades preconfiguradas a las que los clientes puedan suscribirse, eligiendo las funciones que deseen incluir para resolver sus necesidades y descartando aquellas que no necesiten.

Así, los paquetes pueden variar desde un sencillo entorno que se maneje con el ratón y no requiera ningún tipo de conocimiento o instalación especial por el lado del usuario, hasta el suministro de opciones de infraestructura para desarrollo avanzado. (Interoute, 2013-2015)

b) Software como servicio (SAAS)

SaaS es un modelo de distribución del software que proporciona a los clientes el acceso al mismo a través de la red (generalmente Internet), de manera que les libra del mantenimiento de las aplicaciones, de operaciones técnicas y de soporte. Las aplicaciones distribuidas en la modalidad SaaS pueden llegar a cualquier tipo de empresa sin importar su tamaño o su ubicación geográfica. Se trata de un modelo que une el producto (software) al servicio, para dotar a las empresas de una solución completa que permita optimizar sus costes y sus recursos. (Martín, 2007)

c) Infraestructura como servicio (IAAS)

Los recursos informáticos ofrecidos consisten, en particular, en hardware virtualizado, o, en otras palabras, infraestructura de procesamiento. La definición de IaaS abarca aspectos como el espacio en servidores virtuales, conexiones de red, ancho de banda, direcciones IP y balanceadores de carga. Físicamente, el repertorio de recursos de hardware disponibles procede de multitud de servidores y redes, generalmente distribuidos entre numerosos centros de datos, de cuyo mantenimiento se encarga el proveedor del servicio cloud. El cliente, por su parte, obtiene acceso a los componentes virtualizados para construir con ellos su propia plataforma informática. (Interoute, 2013-2015)

1.3.7.6 Arquitectura

Estructura de cloud computing consiste de un conjunto de capas que se encuentran acopladas entre sí para brindar la funcionalidad del sistema, en este caso la arquitectura de Cloud Computing es similar a la arquitectura de red, desde un nivel físico hasta un nivel de aplicación. Esto debido a que Cloud Computing utiliza protocolos similares a los que se usan en Internet como medio de comunicación, ya sea basado en Web o no basado en Web. La arquitectura genérica de Cloud Computing tiene las siguientes capas mencionadas de abajo hacia arriba:

- **Recursos físicos:** incluyen elementos como servidores, almacenamiento y red.
- **Virtualización:** incluye infraestructura virtual como un servicio.
- **Infraestructura:** incluye software de plataforma como servicio.
- **Plataforma:** incluye componentes de aplicación como servicio.
- **Aplicación:** incluye servicios basados en Web y software como servicio.

(Kezherashvili, Computación en la Nube , 2011)

1.3.7.7 Plataformas Cloud

a) Microsoft Azure

La plataforma en la nube de Microsoft es una colección creciente de servicios integrados (proceso, almacenamiento, datos, red y aplicaciones) que le permiten actuar más rápido, llegar más alto y ahorrar dinero. (Microsoft, 2015)

b) AWS

Amazon Web Services (AWS) ofrece soluciones fiables basadas en la nube para atender las necesidades de su empresa. Ejecutar sus soluciones en la nube de AWS puede ayudarle a poner en marcha y

ejecutar sus aplicaciones en menos tiempo, al tiempo que podrá disfrutar de la misma seguridad en la que confían entidades como Pfizer, Intuit y la armada estadounidense (US Navy). AWS también ofrece recursos en todo el mundo para que usted pueda implementar sus soluciones donde estén sus clientes. Con la nube de AWS, es fácil disponer de una amplia gama de servicios, socios y opciones de soporte para que usted pueda centrarse en conseguir que su solución sea todo un éxito. (Amazon Web Services, Inc., 2015)

c) **Google Cloud Platform**

Google Cloud Platform enables developers to build, test and deploy applications on Google's highly-scalable and reliable infrastructure. Choose from computing, storage and application services for your web, mobile and backend solutions. (Google, 2015)

Que traducido significa: “Google Cloud Platform permite a los desarrolladores construir, probar y desplegar aplicaciones en la infraestructura altamente escalable y confiable de Google. Elija de computación, almacenamiento y servicios de aplicaciones para sus soluciones web, móviles y de back-end”.

- **Google App Engine (PAAS)**

Google App Engine is a Platform as a Service (PaaS) offering that lets you build and run applications on Google's infrastructure. App Engine applications are easy to build, easy to maintain, and easy to scale as your traffic and data storage needs change. With App Engine, there are no servers for you to maintain. You simply upload your application and it's ready to go. Que traducido significa: “Google App Engine es una plataforma como servicio (PaaS) que le permite crear y ejecutar aplicaciones en la infraestructura de Google. Aplicaciones App Engine son fáciles de construir, fácil de mantener y fácil de ampliar a medida que cambien sus necesidades

de tráfico y almacenamiento de datos. Con App Engine, no hay servidores para su mantenimiento. Sólo tiene que cargar su aplicación y ya está listo para empezar” (Google, 2015)

- **Google Compute Engine (IAAS)**

Google Compute Engine is a service that provides virtual machines that run on Google infrastructure. Google Compute Engine offers scale, performance, and value that allows you to easily launch large compute clusters on Google's infrastructure. There are no upfront investments and you can run thousands of virtual CPUs on a system that has been designed from the ground up to be fast, and to offer strong consistency of performance.

Traducido es: “Google Compute Engine es un servicio que ofrece máquinas virtuales que se ejecutan en la infraestructura de Google. Google Compute Engine ofrece escala, rendimiento y valor que le permite lanzar fácilmente grandes torres de ordenadores en la infraestructura de Google. No hay inversiones iniciales y se puede ejecutar miles de CPUs virtuales en un sistema que ha sido diseñado desde el principio para ser rápido, y para ofrecer una fuerte consistencia de rendimiento.” (Google, 2015)

d) IBM SmartCloud

IBM SmartCloud proporciona un acceso rápido y flexible a entornos empresariales de servidores virtualizados altamente seguros y es ideal tanto para equipos de desarrollo como de TI ya que proporciona servicios, sistemas y software que permiten satisfacer todas las necesidades corporativas.

Esta línea de soluciones que IBM incluidas en IBM SmartCloud permite a las empresas disponer de un entorno cloud integrado y completo. Además, reduce los costes de operación hasta el 50%. (Cabezudo, 2012)

e) Heroku

Heroku ofrece un servicio PaaS (Plataform as a Service) en donde actualmente se pueden desplegar aplicaciones desarrolladas en Ruby, Node.js, Java, Python, Clojure y Scala. En Heroku definen una aplicación como el código fuente de la aplicación junto a sus dependencias, indicadas, por ejemplo, en el Gemfile de Ruby o el pom.xml de Maven. (GZM, 2015)

1.3.7.8 Seguridad en la Nube

Cierta cantidad de usuarios o empresarios sienten temor cuando les hablan de servicios en la nube, aplicaciones en internet, etc. Lo primero que piensan es en su información y la seguridad, prefieren que todo esté ahí mismo empapelado o guardado en unidades dentro del edificio, pero no se dan cuenta que ello corre mucho más riesgo.

Los datos, cuando están en aplicaciones en cloud, se alojan en DATA CENTERS, empresas específicamente dedicadas a la custodia y salvaguarda de datos de empresas de todo tipo: bancos, entidades financieras, gobierno, multinacionales, pymes, personas como vosotros o como yo. Son empresas que cuentan con todas las medidas de seguridad necesarias, tanto físicas como de software, de forma que no haya jamás una pérdida de información ni de integridad de los datos. La única precaución que hay que tener, es encontrar un DATA CENTER o proveedor que nos dé garantías y prestaciones adecuadas al “valor” que damos a nuestros datos. Y eso, es algo que ya hacemos a día de hoy con otros servicios como Internet, teléfono, etc. (Kezherashvili, Administración de Sistemas operativos, 2011)

En cualquier decisión que se tome siempre habrá riesgos, pero se debe tomar en cuenta simplemente a que empresa estamos contratando y firmar ciertos acuerdos para la confiabilidad y seguridad de nuestra información. Más que desventajas, existen ventajas al contar con la oferta de una empresa dedicada a la custodia de nuestros datos, por ejemplo: Microsoft, cuenta con alrededor de 7 datacenters en el mundo, podría haber un desastre natural en este continente y si tenemos guardada la información en Europa, nuestra información seguiría impecable, sin contar que sigue la norma ISO 27018, respecto a seguridad de nuestra información, en las que aseguran la salvaguarda de ésta y su confiabilidad.

Es la provisión de servicios de seguridad desde la nube, lo que permite que algunos servicios y tecnologías que tradicionalmente solo estaban al alcance de empresas grandes, estén accesibles a Pymes y Administraciones públicas, con las mismas, o incluso a veces más necesidades en este aspecto. Otras de las ventajas que aporta son:

- Provisión del servicio de una manera más rápida, ágil y mucho más fácil de adaptar al crecimiento (escalable), a veces incluso de una manera automatizada.
- Al tratarse de servicios en modo pago por uso permite reducir el presupuesto de inversión en materia de tecnologías y recursos dedicados a la seguridad.
- La gestión de las actualizaciones de las tecnologías de seguridad (antivirus, firewalls, anti-spam, sistemas operativos, servidores web, firmas de los IPS, etc.) necesarias para mantener el nivel de riesgo aceptable, corre a cargo del proveedor de servicios. De esta manera se disminuyen los costes de mantenimiento y el tiempo a dedicar por personal especializado. Estas tecnologías evolucionan cada vez más rápido (para contrarrestar el incremento de amenazas y vulnerabilidades en Internet) y por lo tanto tienen un ciclo de vida

más corto, lo que implica un gasto continuo en tecnologías y procesos, o una disminución de la seguridad (nivel de riesgo) en caso de no hacerlo. (SVT Cloud, 2012)

1.3.7.9 Ventajas

- Entre las ventajas más importantes que ofrecen los WebServices se pueden citar:
- Ofrecen una "tecnología distribuida de componentes" optimizada.
- Evitan los problemas inherentes a la existencia de firewalls, ya que SOAP utiliza HTTP como protocolo de comunicación.
- Permiten una invocación sencilla de métodos, mediante SOAP los clientes o "consumidores de servicios" pueden estar en cualquier plataforma, basta con que soporten XML/SOAP, incluso puede sustituirse SOAP por HTTP.
- Permiten centralizar los datos, independientemente de si los WebServices están distribuidos o no.
- (Rodríguez M. B.)

1.3.7.10 Desventajas

Entre las desventajas más importantes que ofrecen los WebServices se pueden citar:

- Su rendimiento es bajo comparado con otros sistemas como CORBA, DCOM o RMI, especialmente por el uso de protocolos y estándares basados en texto.

- Existe poca información de servicios web para algunos lenguajes de programación.
- Para realizar transacciones no pueden compararse en su grado de desarrollo con los estándares abiertos de computación distribuida como CORBA.

1.3.7.11 Estándares

Los Web Services están contruidos con varias tecnologías que trabajan conjuntamente con los estándares que están emergiendo para asegurar la seguridad y operatibilidad, de modo de hacer realidad que el uso combinado de varios Web Services, independiente de la o las empresas que los proveen, este garantizado. A continuación se describen brevemente los estándares que están ocupando los Web Services. (Saffirio, 2006)

XML

Abreviación de Extensible Markup Language. El XML es una especificación desarrollada por World Wide Web. Permite a los desarrolladores crear sus propias etiquetas, que les permiten habilitar definiciones, transmisiones, validaciones, e interpretación de los datos entre aplicaciones y entre organizaciones. (Saffirio, 2006)

SOAP

Abreviación de Simple Object Access Protocol, es un protocolo de mensajería construido en XML que se usa para codificar información de los requerimientos de los Web Services y para responder los mensajes “antes de enviarlos por la red. Los mensajes SOAP son independientes de los sistemas operativos y pueden ser transportados por los protocolos que funcionan en la Internet, como ser: SMTP, MIME y HTTP. (Saffirio, 2006)

WSDL

Abreviación de Web Services Description Language, es un lenguaje especificado en XML que se ocupa para definir los Web Service como colecciones de punto de comunicación capaces de intercambiar mensajes. El WSDL es parte integral de UDDI y parte del registro global de XML, en otras palabras es un estándar de uso público (no se requiere pagar licencias ni royalties para usarlo). (Saffirio, 2006)

UDDI

Abreviación de Universal Description, Discovery and Integration. Es un directorio distribuido que opera en la Web que permite a las empresas publicar sus Web Services, para que otras empresas conozcan y utilicen los Web Services que publican, opera de manera análoga a las páginas amarillas. (Saffirio, 2006)

1.3.8 Arquitectura de un Web Service.

La arquitectura básica del modelo de web services describe a un consumidor, un proveedor y ocasionalmente un corredor (broker). Relacionados con estos agentes están las operaciones de publicar, encontrar y enlazar. (GXportal, 2013)

La idea básica consiste en que un proveedor publica su servicios en un corredor, luego un consumidor se conecta el corredor para encontrar los servicios deseados y una vez que lo hace se realiza un lazo entre el consumidor y el proveedor. (GXportal, 2013)

Cada entidad puede jugar alguno o todos los roles.

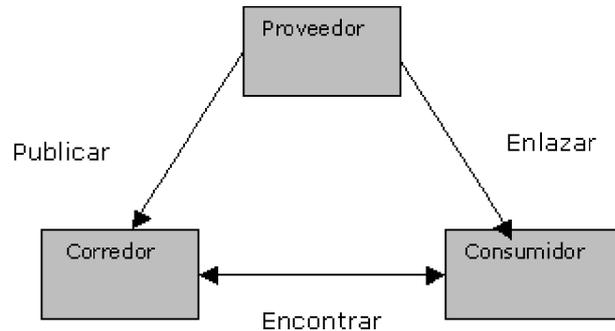


Ilustración 2 Arquitectura web services

Por todo lo anterior hay ciertos requerimientos a la hora de desarrollar o consumir un web service:

- Una forma estándar de representar los datos.
- XML es la opción obvia para este requerimiento.
- Un formato común y extensible de mensajes.
- SOAP es el elegido en este caso; SOAP es un protocolo liviano para el intercambio de información. Más adelante en este documento lo veremos con más detalle.
- Un lenguaje común y extensible para describir los servicios.
- La opción en este caso es WSDL. Es un lenguaje basado en XML desarrollado en forma conjunta por IBM y Microsoft. Lo veremos con más detalle más adelante en este documento.
- Una forma de descubrir los servicios en Internet.
- UDDI se utiliza en este caso; el mismo especifica un mecanismo para publicar y localizar los servicios por parte de los proveedores y consumidores respectivamente. Se verá con más detalle más adelante en este documento. (GXportal, 2013)

1.3.9 Infraestructura de TI.

La infraestructura de TI consiste en un conjunto de dispositivos físicos y aplicaciones de software que se requieren para operar toda la empresa. Sin embargo, la infraestructura de TI también es un conjunto de servicios a lo largo y ancho de la empresa, presupuestados por la administración y que abarcan capacidades tanto humanas como técnicas.

Los servicios que una empresa es capaz de proveer a sus clientes, proveedores y empleados son una función directa de su infraestructura de TI. En un plano ideal, esta infraestructura debería apoyar la estrategia de negocios y sistemas de información de la empresa. Las nuevas tecnologías de información tienen un potente impacto en las estrategias de negocios y de TI, así como en los servicios que se pueden ofrecer a los clientes. (Silvestrucci, 2013)

- Plataformas de cómputo que se utilizan para proporcionar servicios de cómputo que conectan a empleados, clientes y proveedores dentro de un entorno digital coherente, el cual incluye grandes mainframes, computadoras de escritorio y portátiles, así como asistentes digitales personales (PDAs) y dispositivos para Internet.
- Servicios de telecomunicaciones que proporcionan conectividad de datos, voz y video a empleados, clientes y proveedores.
- Servicios de administración de datos que almacenan y manejan datos corporativos y proveen capacidades para analizar los datos.
- Servicios de software de aplicaciones que proporcionan capacidades a toda la empresa, como sistemas de planeación de recursos empresariales, de administración de las relaciones con el cliente, de administración de la cadena de suministro y de administración del conocimiento, los cuales son compartidos por todas las unidades de negocios.

- Servicios de administración de instalaciones físicas que desarrollan y manejan las instalaciones físicas requeridas por los servicios de cómputo, de telecomunicaciones y de administración de datos.
- Servicios de administración de TI que planean y desarrollan la infraestructura, coordinan los servicios de TI entre las unidades de negocios, manejan la contabilidad de los gastos en TI y proporcionan servicios de administración de proyectos.
- Servicios de estándares de TI que dotan a la empresa y sus unidades de negocios de las políticas que determinan cuál tecnología de información se utilizará, en qué momento y de qué manera.
- Servicios de entrenamiento en TI que proporcionan a los empleados capacitación en el uso de los sistemas y a los gerentes, capacitación sobre la manera de planificar y manejar las inversiones en TI.
- Servicios de investigación y desarrollo de TI que proporcionan a la empresa investigación sobre proyectos e inversiones de TI potenciales que podrían ayudar a la empresa a diferenciarse en el mercado. (Silvestrucci, 2013)

1.3.9.1 Seguridad

La seguridad es un proceso de dos etapas cuyo objetivo es evitar que un usuario acceda a un recurso para el que no debería tener acceso.

En el primer paso del proceso, el sistema de seguridad identifica quién es el usuario obligándolo a enviar algún tipo de identificación. Esto se llama autenticación, y significa que el sistema está tratando de averiguar quién eres.

Una vez que el sistema sabe quién eres, el siguiente paso es decidir si deberías tener acceso a un determinado recurso. Esta parte del proceso se llama

autorización, y significa que el sistema está comprobando si tienes suficientes privilegios para realizar una determinada acción. (LibrosWeb.es, 2015)

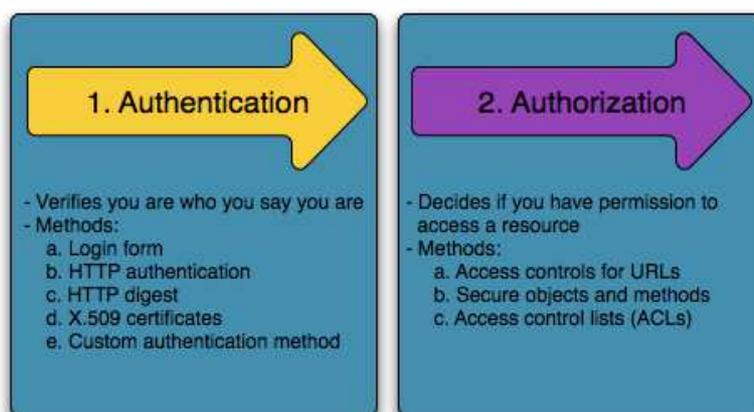


Ilustración 3 Autenticación y Autorización

a) Firewalls (autenticación)

El trabajo del firewall consiste en determinar si el usuario necesita estar autenticado, y si lo necesita, enviar una respuesta al usuario para iniciar el proceso de autenticación.

Un firewall se activa cuando la URL de una petición entrante concuerda con el valor de su opción de configuración pattern. En este ejemplo el valor de pattern (^) concuerda con cualquier petición entrante. No obstante, el hecho de que el firewall esté activado no significa que el navegador muestra la caja de login+contraseña para todas las URL.

Al igual que sucedía anteriormente, cuando el usuario realiza su petición, el firewall no solicita ningún tipo de identificación. Sin embargo, en cuanto la capa de control de acceso deniega el acceso al usuario (porque los usuarios anónimos

no cuentan con el rol ROLE_ADMIN), el firewall toma el control de la aplicación e inicia el proceso de autenticación.

La autenticación se puede hacer de varias maneras: a través de un formulario de acceso, con la autenticación básica de HTTP, e incluso a través de Twitter o Facebook. (LibrosWeb.es, 2015)

a. Autenticación multi-factor

La autenticación multifactor (MFA) es un sistema de seguridad que requiere más de una forma de autenticación para verificar la legitimidad de una transacción. Combina dos o más credenciales independientes: lo que sabe el usuario (contraseña), lo que tiene el usuario (token de seguridad) y lo que es el usuario (verificación biométrica). (Rouse, 2014)

b. Control de acceso (autorización)

Una vez que el usuario se ha autenticado, comienza la autorización, que proporciona un mecanismo estándar y muy potente para decidir si un usuario puede acceder a algún recurso (una URL, un objeto, una llamada a un método, entre otros). Su funcionamiento se basa en asignar roles específicos a cada usuario y después hacer que las diferentes partes de la aplicación requieran de diferentes roles para poder acceder.

El proceso de autorización tiene por tanto dos componentes principales:

- El usuario dispone de uno o más roles.
 - Un recurso requiere de un rol específico para poder acceder a él.
- (LibrosWeb.ec, 2015)

1.4 FUNDAMENTACIÓN LEGAL

La fundamentación legal se dirige a la empresa pública, por ser estudiantes de una universidad pública el objeto de muestra.

En base a la LEY ORGÁNICA DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA, considera lo siguiente

Que el artículo 81 de la Constitución Política de la República, (LEY ORGÁNICA DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA) respecto del manejo de la cosa pública y la rendición de cuentas a la que están sujetos todos los funcionarios del Estado, y demás entidades obligadas por esta Ley:

- Que es necesario hacer efectivo el principio de publicidad de los actos, contratos y gestiones de las instituciones del Estado y de aquellas financiadas con recursos públicos o que por su naturaleza sean de interés público.
- Que la misma norma constitucional establece que no existirá reserva respecto de informaciones que reposen en archivos públicos, excepto de aquellas que por seguridad nacional no deben ser dadas a conocer.
- Que la libertad de información está reconocida tanto en el artículo 19 del Pacto Internacional de Derechos Civiles y Políticos, como en el artículo 13 de la Convención Interamericana de Derechos Humanos; y, En ejercicio de sus facultades constitucionales y legales, expide la LEY ORGÁNICA DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA
- Esta ley de la pauta de la necesidad legal de las empresas públicas de tener canales de difusión de su información, y en los tiempos actuales, es inaceptable el considerar una empresa que no tenga canales de difusión en las

redes sociales, más aun cuando está estipulado por la ley el que se realice la difusión de esta información.

(LEY ORGÁNICA DE TRANSPARENCIA Y ACCESO A LA INFORMACIÓN PÚBLICA)

1.5 CONCLUSIONES RELACIONADAS AL MARCO TEÓRICO EN REFERENCIA AL TEMA DE INVESTIGACIÓN

Microsoft Azure es una plataforma que ha ido evolucionando con el paso de los tiempos, junto a otras tecnologías como es el Cloud Computing. Su aporte científico, así como el de otras plataformas Cloud han marcado huellas. La educación necesita fortalecerse con actualización de conocimientos cada día, el usar y experimentar con nuevas herramientas es lo que caracteriza a un ingeniero, es necesario conocer más sobre lo que causa impacto en el entorno informático y sacar el potencial de nuestras ideas y proyectos a través de; así como conocer de funcionamientos y bases sobre las que giran los elementos que usamos para el desarrollo.

CAPITULO II: DIAGNÓSTICO O ESTUDIO DE CAMPO

2.1 INTRODUCCIÓN

Mediante la investigación que se encontrará descrita en este capítulo, se van a explicar los métodos y diferentes herramientas de investigación mediante los cuales se recopilan los datos e información para el desarrollo de este trabajo de titulación.

Para encontrar la situación real del problema planteado, se llevará a cabo el análisis de estos datos obtenidos del objeto de la investigación. En nuestro caso más específico, la descripción del grado de conocimiento de estudiantes de últimos niveles en cuanto a plataformas de servicios, además de entender los diferentes aspectos del Cloud Computing.

2.2 TIPO DE INVESTIGACIÓN

El proceso de investigación requiere precisión y la definición de un modelo de estrategias que sirvan como orientación, por esta razón se usó el tipo de investigación según el nivel de medición: cuantitativa, debido a sus técnicas estadísticas y al uso de sus magnitudes numéricas.

La investigación que sigue una metodología cuantitativa supone un planteamiento, un acercamiento a la realidad objeto de estudio y a la teoría, y unos fines de la investigación característicos:

- En primer lugar, el objeto de análisis es una realidad observable, medible y que se puede percibir de manera precisa.
- La relación entre la teoría y la hipótesis es muy estrecha pues la segunda deriva de la primera. A partir de un marco teórico se formula una hipótesis,

mediante un razonamiento deductivo, que posteriormente se intenta validar empíricamente.

- Se busca establecer una relación de causa-efecto entre dos fenómenos. Dicha relación está ligada con la interconexión entre conceptos que supone la hipótesis. La validación de la hipótesis supone explicitar esa relación de causa-efecto latente en dicha hipótesis
- Se analizan las variables, tratadas con procedimientos matemáticos y estadísticos. Una variable es una característica que puede adoptar distintos valores.
- Tiene capacidad de predicción y generalización. Se trabaja sobre una muestra representativa del universo estudiado.
- (Instituto Cervantes, 1997-2015)

Más específicamente en el ámbito de esta investigación, obtendremos en gran parte del proceso investigativo como resultado principal pruebas de conocimientos básicos sobre Cloud Computing, por ende se utiliza en mayor parte la información del tipo cuantitativo, analizando de manera científica-numérica.

2.3 MÉTODOS DE INVESTIGACIÓN

Se requiere la utilización de un método investigativo la obtención y análisis de información, para lograr los objetivos propuestos anteriormente. Al ser este un estudio estadístico con variables numéricas, se lo ejecutó a través de una suite ofimática, utilizando el método empírico-analítico, y debido al amplio proceso de investigación que requiere el tema y la necesidad de obtener información específica que requiere la consulta de diversas obras de referencia se utilizará también el método bibliográfico.

La investigación empírica se puede definir como "la investigación basada en la experimentación o la observación (evidencias)". Este tipo de investigación es llevada a cabo para poner a prueba una hipótesis.

El tema central en el método científico es que todo aporte debe ser empírico, lo que significa que es basado en la evidencia. Teniendo esto en cuenta, la palabra "empírica" también se refiere a trabajar con hipótesis que pueden comprobarse mediante la observación y los experimentos. Los datos empíricos son producidos por la experiencia y la observación.

El método de investigación bibliográfica es el sistema que se sigue para obtener información contenida en documentos. En sentido más específico, es el conjunto de técnicas y estrategias que se emplean para localizar, identificar y acceder a aquellos documentos que contienen la información pertinente para la investigación.

Pueden elaborarse bibliografías de manera independiente, cuyo objetivo sea la compilación de obras para su uso con cualquier fin y por cualquier investigador. Las bibliografías que resultan de estos procesos, son los frutos más constantes de cualquier labor investigadora. (Prada, 2001)

2.4. HERRAMIENTA(S) DE RECOLECCIÓN DE DATOS

Para el desarrollo de este proyecto fue necesario utilizar herramientas de recolección que nos permitieron alcanzar los objetivos propuestos, se utilizaron las siguientes herramientas:

- Encuestas

Por medio del diseño de un cuestionario de preguntas se logró registrar diferentes opiniones, percepciones y conceptos de estudiantes de últimos niveles

de la carrera de ingeniería de sistemas, respecto al Cloud Computing, plataformas de servicios y su rol como estudiantes y próximos profesionales.

- **Observación.**

Al ser una investigación educativa la observación de campo ha sido indispensable. Se ha llevado a cabo en el centro de estudios, lo que permitió el análisis de comportamientos y conductas de las masas en torno a las nuevas tecnologías, su aprendizaje y motivación respecto a éstas.

- **Análisis Documental**

En la actualidad la documentación existente sobre el Cloud Computing crece como espuma, por ello la web ha aportado manera fundamental al desarrollo cognitivo de la búsqueda, permitiendo dar un aporte más científico y puntual a la hipótesis que queremos probar.

2.5. FUENTES DE INFORMACIÓN DE DATOS

a) Primarias. La información recopilada de las redes sociales, de las revistas informativas y más, sobre conocimientos sobre Cloud Computing y plataformas de servicios, los cursos a los que pertenecen o las comunidades en las que incursionan los distintos estudiantes de últimos niveles, sirven de base para una investigación eficiente respecto a las necesidades que los estudiantes de últimos niveles requieren. Es necesario añadir que, lo antes expuesto es producto de las diversas técnicas de investigación aplicadas, que no hacen más que evidenciar los aspectos importantes que se deben tomar en cuenta.

b) Secundarias. Se utilizan las siguientes técnicas:

a) Encuestas a estudiantes de últimos niveles.

- b) Encuestas a docentes de las áreas técnicas de la Facultad de estudio.
- c) Información extraída del Internet.

Durante la investigación las fuentes principales de información de datos fueron los datos estadísticos que se pudieron obtener de las encuestas suministradas a los estudiantes en las que se investigaba el nivel de conocimiento que tienen sobre el tema, sus aspiraciones y necesidades como informáticos.

2.6. INSTRUMENTAL OPERACIONAL

2.6.1. Estructura y características de lo(s) instrumento(s) de recolección de datos.

La realización del proyecto ha constituido la utilización de uno de los instrumentos más efectivos, como es la encuesta en cuanto a instrumentos se refiere, sin restar importancia a las otras técnicas usadas en este proyecto, cada una con su fin respectivo. La realización de encuestas se efectuaron con el fin de dar mayor credibilidad a la investigación con el factor muestra. Se las elaboraron con una estructura lo bastante objetiva y precisa para obtener resultados puntuales.

Encuesta.

Se estructuraron 3 encuestas, 2 de ellas orientadas a los alumnos de últimos niveles (7mo, 8vo y 9no) de la Facultad de Ciencias Informáticas de la ULEAM de Manta, y la tercera a los docentes de las áreas técnicas de la misma. El objetivo de estas encuestas fue medir el conocimiento en general sobre Inteligencia Artificial, Aprendizaje automático, Computación en la Nube y Plataformas de Servicios; y de los estudiantes en específico también analizar el

interés por las nuevas tecnologías, el desarrollo de software, el emprendimiento y la obtención de experiencia.

2.7. ESTRATEGIA OPERACIONAL PARA LA RECOLECCIÓN Y TABULACIÓN DE DATOS

OBSERVACIÓN

En base a la experimentación y lógica empírica se realiza el análisis de los comportamientos de los estudiantes informáticos en el centro de estudio, así como en las bibliotecas, lo que permitirá llegar a deducciones por medio de patrones para la presentación de la solución.

ANÁLISIS DOCUMENTAL

La recolección de información a través de los diversos métodos y el análisis de requerimientos y necesidades, se analizarán para efectuar una solución con un impacto significativo, que permita cambiar o mejorar la actual situación.

2.7.1. Plan de recolección, tabulación, análisis e interpretación de los datos.

La recolección de los datos fue realizada en Manta, ciudad de Ecuador en la Universidad Laica Eloy Alfaro de Manabí, a los estudiantes entre 7mo y 9no nivel, y a docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la ULEAM, al ser individuos experimentados con mayores necesidades en un mundo tecnológico, y al servicio de la educación según corresponde.

Al culminar las encuestas, se procedió a la tabulación mediante herramientas informáticas y gráficos estadísticos, para facilitar el proceso y poder presentar y analizar los datos de manera efectiva para una investigación óptima.

Se usó el programa Microsoft Excel 2013 para la tabulación, lo que ayudó al cálculo y la creación de gráficos que nos permitirán obtener una mejor visión de lo que se busca. Todos los datos se han obtenido por medio de las encuestas realizadas a los distintos estudiantes dentro del rango de 7mo a 9no nivel, y a profesores de las áreas técnicas de la facultad.

2.8. PLAN DE MUESTREO

El muestreo de la investigación es el proceso mediante el cual se toma una muestra de la población, y en base a esta se procede a realizar las técnicas de investigación las cuales nos permitirán estudiar a la población de una manera eficiente.



Ilustración 4 Plan de muestreo

La segmentación se realizó en base a los docentes y estudiantes que existen en la Universidad Laica Eloy Alfaro de Manabí, específicamente a los

estudiantes entre 7mo, 8vo y 9no nivel y a docentes de las áreas técnicas de la Facultad de Ciencias Informáticas. Se escogieron estos dos segmentos debido a que los estudiantes de último nivel cuentan con más experiencias, tienen mayores conocimientos y necesidades en sus proyectos estudiantiles como pre-profesionales; y los docentes por su parte, cuentan con mayores conocimientos sobre nuevas herramientas, tecnologías y métodos, además de estar al servicio y ayuda de los alumnos.

Para la determinación de la muestra no se utilizó técnica alguna. Al contar con una población pequeña, se procedió a encuestar a todos los estudiantes, con el fin de obtener información más precisa y por lo tanto una hipótesis más acertada. La muestra considerada para aplicar la encuesta es de **106 estudiantes y 14 docentes**.

2.9. PRESENTACIÓN Y ANÁLISIS DE LOS RESULTADOS

2.9.1. Presentación y descripción de los resultados obtenidos.

Los datos obtenidos después de ser tabulados y analizados presentan la siguiente representación en un gráfico de tipo pastel donde se muestra de una manera fácil los resultados de las encuestas que posteriormente servirán para resolver un análisis para cada respectiva pregunta:

Encuesta realizada a los estudiantes de últimos niveles (7mo, 8vo, 9no) de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Pregunta No. 1: Para usted Cloud Computing es:

Cuadro #1

Orden	Alternativas	f	%
1	Un paradigma que permite ofrecer servicios de computación a través de Internet.	92	87%
2	El uso de equipos de telecomunicaciones y computadoras para la transmisión, el procesamiento y el almacenamiento de datos.	12	11%
3	Una ciencia que estudia la creación y diseño de entidades capaces de resolver problemas cotidianos por sí mismas	0	0%
4	No sé	2	2%
TOTAL		106	100%

Fuente: Estudiantes de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #1

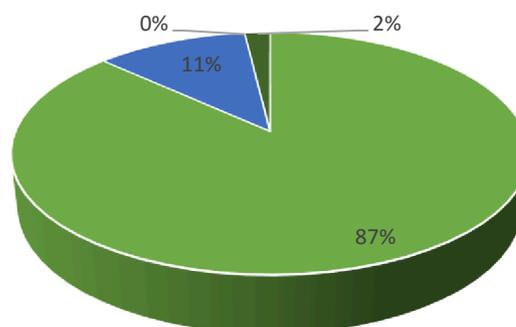


Ilustración 5 Encuesta 1 - Pregunta 1

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 87% conoce que es el Cloud Computing, y un 11% desconoce, sosteniendo un concepto básico para el entendimiento de Machine Learning en Azure.

Pregunta No. 2: Para usted PAAS es:

Cuadro #2

Orden	Alternativas	f	%
1	Plataforma como Servicio	52	51%
2	Puertos como Servicio	3	1%
3	Portal como Servicio	2	2%
4	No sé	49	46%
TOTAL		106	100%

Fuente: Estudiantes de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #2

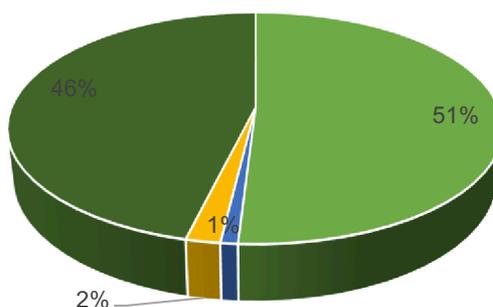


Ilustración 6 Encuesta 1 - Pregunta 2

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 51% conoce sobre Plataformas como servicios, y un 49% de ellos desconoce. Esta capa es uno de los pilares del cloud computing, y sobre esta corre el servicio de Machine Learning, cabe recalcar que aún sin saberlo se puede usar el servicio, pero como buen informático debemos saber cómo funciona, si no es específicamente porque es muy amplio, si de forma transparente.

Pregunta No.3. Para usted SAAS es:

Cuadro #3

Orden	Alternativas	F	%
1	Shareware como Servicio	5	5%
2	Socket como Servicio	0	0%
3	Software como Servicio	50	50%
4	No sé	45	45%
TOTAL		100	100%

Fuente: Estudiantes de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #3

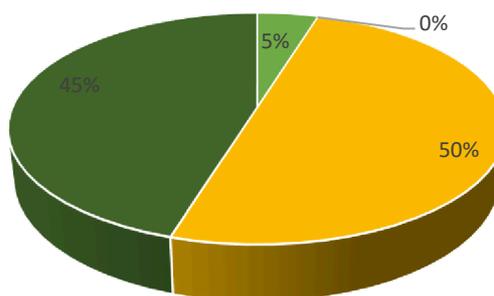


Ilustración 7 Encuesta 1 - Pregunta 3

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 50% conoce sobre Software como Servicio, y un 48% desconoce. Esta capa es uno de los pilares del cloud computing, aunque no es de vital importancia para la investigación, el que tengan un concepto básico se prefiere, para que puedan abrir horizontes en Azure.

Pregunta No.4. Para usted IAAS es:

Cuadro # 4

Orden	Alternativas	F	%
1	Infraestructura como Servicio	49	46%
2	Interfaz como Servicio	6	6%
3	Inteligencia como Servicio	1	1%
4	No sé	50	47%
TOTAL		106	100%

Fuente: Estudiantes de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #4

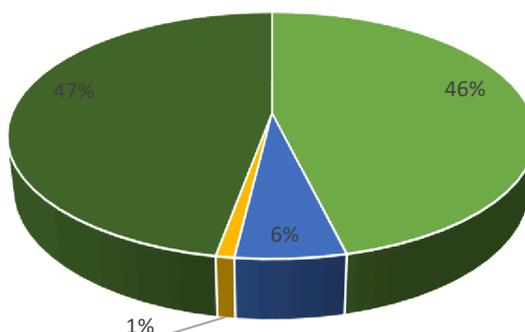


Ilustración 8 Encuesta 1 - Pregunta 4

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 46% conoce que es Infraestructura como servicio, y un 54% desconoce. Esta capa es uno de los pilares del cloud computing, aunque no es de vital importancia para la investigación, el que tengan un concepto básico se prefiere, para que puedan abrir horizontes en Azure

Pregunta No. 5. ¿Conoce los tipos más conocidos de nube existentes en la actualidad? Encierre la opción escogida.

Cuadro #5

Orden	Alternativas	f	%
1	Nube rápida	0	0%
2	Nube privada	84	38%
3	Nube comunitaria	8	4%
4	Nube distribuida	16	7%
5	Nube pública	68	31%
6	Nube híbrida	40	18%
7	Nube solitaria	0	0%
8	No sé	4	2%
TOTAL		220	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #5

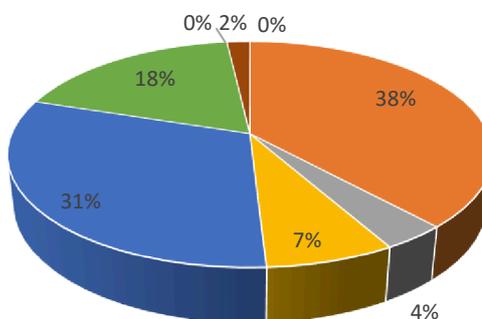


Ilustración 9 Encuesta 1 - Pregunta 5

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 40% conoce sobre los 3 tipos de nubes principales y más conocidas en la actualidad, un 44% conoce sobre alguno de ellos, un 7% sobre otro tipo de nubes y un 6% desconoce totalmente. Aunque no es obligatorio para el uso de Machine Learning, con esta pregunta se buscaba obtener un nivel cercano al conocimiento que tienen sobre cloud computing y plataformas de servicios, para conocer la partida de la propuesta.

Pregunta No.6. ¿Ha escuchado sobre alguna de las siguientes plataformas de servicios? Encierre la opción escogida.

Cuadro #6

Orden	Alternativas	F	%
1	MWA	69	50%
2	AWS	8	6%
3	Google Cloud Platform	52	38%
4	IBM SmartCloud	3	2%
5	Heroku	3	2%
6	No sé	2	2%
TOTAL		135	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #6

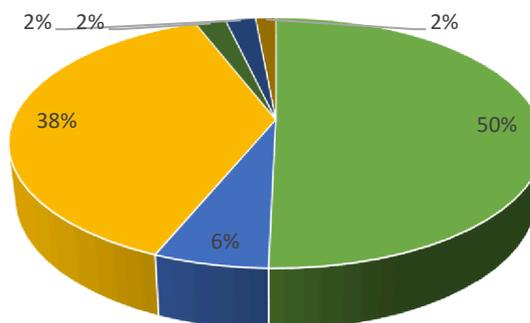


Ilustración 10 Encuesta 1 - Pregunta 6

Análisis: Según encuesta realizada a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 50% ha escuchado sobre Microsoft Azure, el 38% sobre otras plataformas de servicios, y un 2% nunca ha escuchado, lo que le da a la plataforma de servicio que se usa en el proyecto una ventaja, ante cualquier otra, siendo la más propensa a llamar la atención, tomando en cuenta además que es la mejor opción en cuanto a aprendizaje por tener la interfaz más intuitiva.

Pregunta No. 7. Para usted Machine Learning es:

Cuadro #7

Orden	Alternativas	F	%
1	Rama de la inteligencia artificial que busca desarrollar técnicas para que las computadoras aprendan.	44	36%
2	Razonamiento mediante lógica formal análogo al pensamiento abstracto inhumano.	8	4%
3	No sé	54	60%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #7

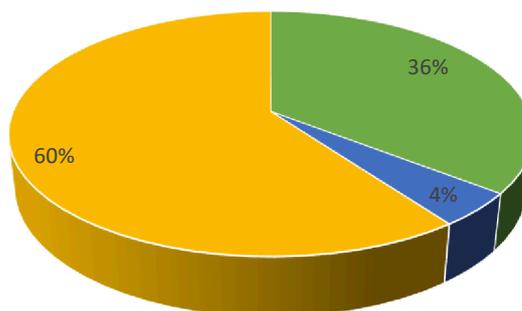


Ilustración 11 Encuesta 1 - Pregunta 7

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, un 36% conoce sobre Machine Learning y el otro 60% desconoce. Se debe al bajo protagonismo que tuvo durante mucho tiempo, y la vida que recién está tomando con las nuevas tecnologías; sólo ayuda a determinar que no existe gran conocimiento, soportando la investigación principal del estudio y determinando el punto de partida de la propuesta.

Pregunta No. 8. Sabe usted, ¿Cuál de las siguientes alternativas son tipos de Machine Learning? Encierre la opción escogida.

Cuadro #8

Orden	Alternativas	F	%
1	Clustering	1	1%
2	Supervisado	1	1%
3	Parallel	1	1%
4	No supervisado	1	1%
5	No sé	102	96%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #8

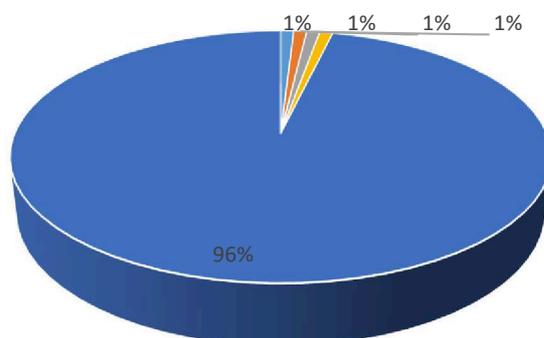


Ilustración 12 Encuesta 1 - Pregunta 8

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 2% conoce los tipos de aprendizaje que tiene el Machine Learning y el 98% desconoce. Se buscaba medir el nivel de conocimiento que tenía el 36% de alumnos que ya conocía o había escuchado sobre Machine Learning según la pregunta anterior (Pregunta No. 7); entonces se puede concluir que los que han escuchado o conocen, tienen un nivel bajo.

Pregunta No. 9. ¿Bajo qué plataformas desarrolla aplicaciones móviles en la actualidad?

Cuadro #9

Orden	Alternativas	f	%
1	Android	37	33%
2	IOS	5	4%
3	Windows Phone	18	16%
4	No desarrollo aplicaciones móviles en la actualidad	54	47%
TOTAL		114	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #9

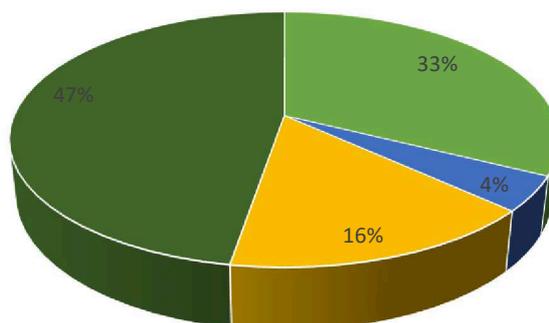


Ilustración 13 Encuesta 1 - Pregunta 9

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 53% desarrolla aplicaciones móviles, y el 47% restante no lo hace, por lo que éste 53% saldrían mayormente beneficiados con el estudio, al no sólo tener al alcance conocimiento, sino una herramienta para darle mayor valor a lo que hacen.

Pregunta No. 10. ¿Qué prefiere usted?

Cuadro #10

Orden	Alternativas	F	%
1	Crear una aplicación web con herramientas como Wordpress	85	82%
2	Crear una aplicación web desarrollando en lenguajes como php, c#, entre otros	12	13%
3	Ninguna de las anteriores	9	5%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #10

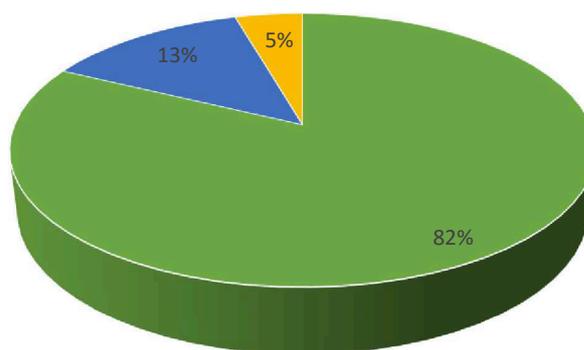


Ilustración 14 Encuesta 1 - Pregunta 10

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 9% prefiere no hacer aplicaciones, el 80% prefiere trabajar con herramientas ya personalizadas para construir aplicaciones webs y un 11% prefiere los lenguajes de programación, haciendo un 91% de estudiantes potenciales en el desarrollo de software, y por lo tanto con mayores ventajas si aprenden sobre aprendizaje automático.

Encuesta realizada a los estudiantes de últimos niveles (7mo, 8vo, 9no) de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Pregunta No.1. ¿Qué tan frecuente es la investigación que realiza en su casa respecto a lo impartido en clases?

Cuadro #1

Orden	Alternativas	F	%
1	Todos los días	56	53%
2	3 días por semana	32	30%
3	1 día por semana	15	14%
4	Ningún día	3	3%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #1

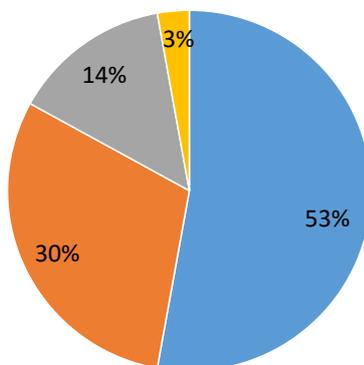


Ilustración 15 Encuesta 2 – Pregunta 1

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 53% investigan todos los días sobre temas relacionados con la carrera, un 30% frecuentemente, dejando un 18% que lo hacen una vez a la semana o pocas veces al mes, demostrando que la tasa de investigación en los alumnos es alta, y que tal vez no son especialistas, pero si tienen abiertos sus horizontes, y conocimientos necesarios.

Pregunta No. 2. Al terminar su carrera como Ingeniero de Sistemas, le gustaría:

Cuadro #2

Orden	Alternativas	f	%
1	Hacer una maestría, postgrado o especialización	66	62%
2	Hacer otra carrera	9	8%
3	Hacer cursos cortos sobre informática	23	22%
4	Ninguna de las anteriores	8	8%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #2

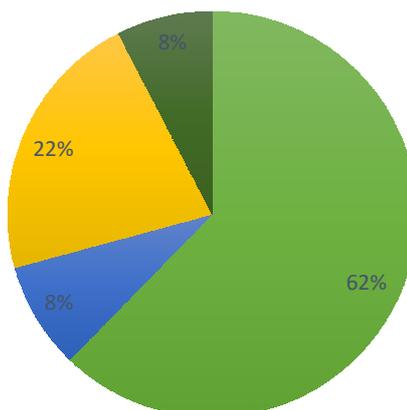


Ilustración 16 Encuesta 2 - Pregunta 2

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, un 84% aspiran seguir estudiando, preparándose e investigando en el campo de la informática, un 8% quiere hacer otra carrera y un 8% aún no sabe lo que quiere hacer. En base a los resultados se puede notar el valor que le dan a la educación y la mentalidad con la que ya cuentan, lo que hace fácil la socialización de la propuesta, al tratar con adultos que quieren seguir desarrollándose.

Pregunta No.3. Si hubiera una tecnología que le diera valor a sus aplicaciones, la aprendería.

Cuadro #3

Orden	Alternativas	f	%
1	Si me incursionaría en el reto.	61	57%
2	Prefiero seguir usando las tecnologías que ya conozco.	4	4%
3	No desarrollo aplicaciones.	35	33%
4	No sé.	6	6%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #3

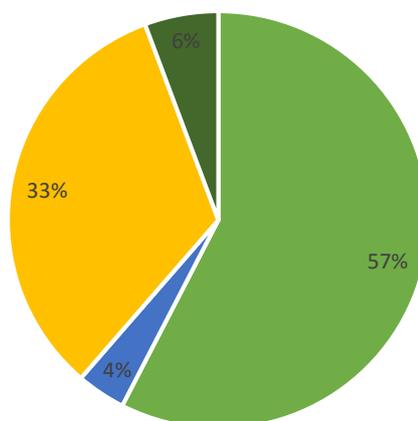


Ilustración 17 Encuesta 2 - Pregunta 3

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, un 68% se incursionaría a aprender nuevas tecnologías, el 6% prefiere seguir usando tecnologías que conocía con anterioridad y un 26% deja abierta la oportunidad, si el resultado es poder aumentar el valor de nuestro trabajo.

Pregunta No. 4. ¿Cuáles son las áreas en las que más destaca? Encierre la opción escogida.

Cuadro #4

Orden	Alternativas	f	%
1	Desarrollo de Software	44	32%
2	Ingeniería de Software	20	15%
3	Admin. de Base de Datos	32	23%
4	Inteligencia Artificial	4	3%
5	Infraestructura en Redes	28	21%
6	Electrónica y Digitales	4	3%
7	No sé	4	3%
TOTAL		136	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #4

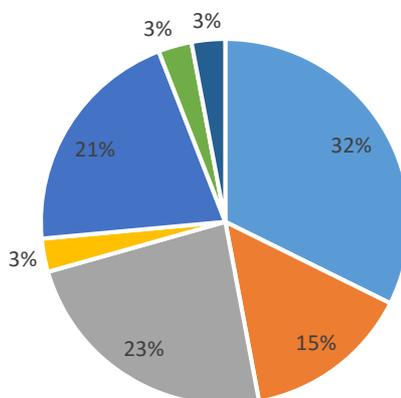


Ilustración 18 Encuesta 2 - Pregunta 4

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, un 73% tiene más afinidad por el desarrollo de software, Administración de BD, Ing. De Software e Inteligencia Artificial dándonos una idea del porcentaje que podría sacar más beneficios, sin descartar que el ML se lo puede aplicar en cualquier campo siempre y cuando tenga aplicación.

Pregunta No. 5. Marque con X los dispositivos que posee:

Cuadro #5

Orden	Alternativas	f	%
1	Celular básico o de gama baja	44	19%
2	Smartphone	62	26%
3	Tablet	11	5%
4	Laptop	86	36%
5	Desktop	32	14%
6	Wearables	0	0%
7	Ninguno de los anteriores	0	0%
TOTAL		209	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #5

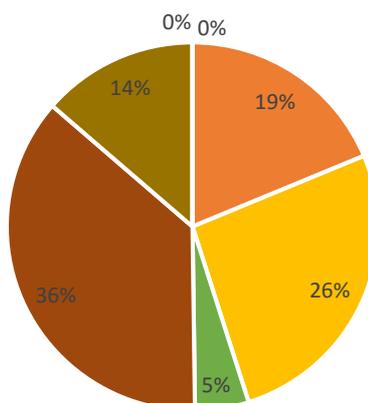


Ilustración 19 Encuesta 2 - Pregunta 5

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 36% tienen laptops y el 14% tiene desktops lo que indica que por lo menos la mitad de ellos poseen herramientas para poder llevar a cabo experimentos de ML desde casa, y el 45%, en la facultad por medio de sus laboratorios, siempre y cuando se cuente con internet en ambos; señalando que los medios principales están al alcance si se quiere ir más a fondo o poner en práctica.

Pregunta No. 6. ¿Bajo qué denominación calificaría el acceso a Internet en general tomando en cuenta el de su hogar si lo tuviera, el de la facultad, o ambos?

Cuadro #6

Orden	Alternativas	f	%
1	Alta disponibilidad	64	60%
2	Nivel medio	35	33%
3	Nivel bajo o casi nulo	7	7%
TOTAL		106	100%

Fuente: Estudiante de 7mo, 8vo y 9no Nivel de Ing. De Sistemas

Elaborada por: Verónica E. Muentes Molina

Gráfico #6

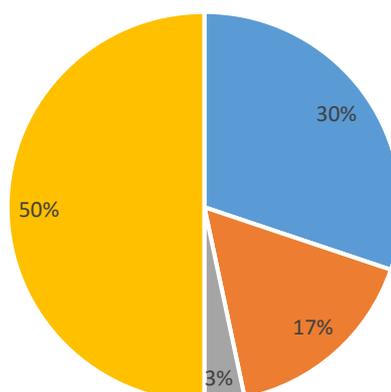


Ilustración 20 Encuesta 2 - Pregunta 6

Análisis: Según encuestas realizadas a estudiantes de últimos niveles de la Facultad de Ciencias Informáticas de la ULEAM, el 60% tiene alta disponibilidad al internet, el 33% en un nivel medio y el 7% un nivel bajo, lo que indica que el 93% cuentan con un internet aceptable para poder hacer experimentos y hacer uso de la plataforma Azure.

Encuesta realizada a los docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Pregunta No. 1 Para usted el aprendizaje es:

Cuadro # 1

Orden	Alternativas	f	%
1	Adquisición del conocimiento de algo por medio del estudio, el ejercicio o la experiencia, en especial de los conocimientos necesarios para aprender algún arte u oficio.	9	64%
2	Facultad del ser humano para comprender por medio de la razón la naturaleza, cualidades y relaciones de las cosas.	4	29%
3	No sé.	1	7%
TOTAL		14	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 1

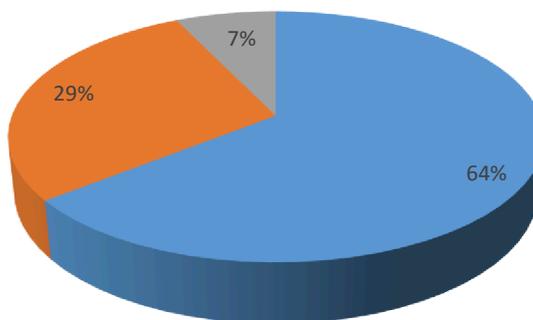


Ilustración 21 Encuesta 3 - Pregunta 1

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, un 64% saben lo que significa el aprendizaje, un 29% lo confunden con conocimiento y un 7% que representa a uno, desconoce su significado; los resultados determinan que se ha hallado un punto de partida en la investigación.

Pregunta No. 2 Si Machine Learning o Aprendizaje de Máquina es una rama de la Inteligencia Artificial, que desarrolla técnicas que permiten aprendizaje de las máquinas para hacer predicciones, diagnósticos, detecciones, etc. En la escala del 1 como bajo y 5 como máximo, determine la importancia, que cree usted tiene el aprendizaje del Machine Learning en la carrera de Ingeniería de Sistemas.

Cuadro #2

Orden	Alternativas	f	%
1	1	0	0%
2	2	0	0%
3	3	3	21%
4	4	7	50%
5	5	4	29%
TOTAL		14	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 2

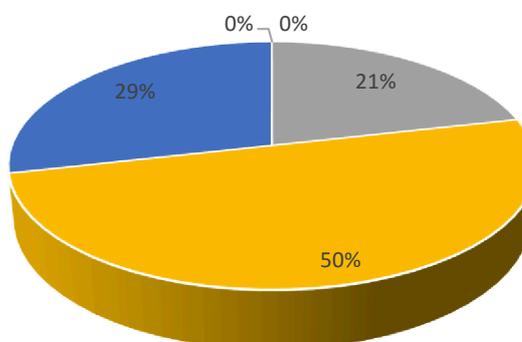


Ilustración 22 Encuesta 3 - Pregunta2

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, un 79% consideró muy importante conocer sobre Machine Learning, y un 21% que era lo suficientemente importante; los resultados obtenidos fueron positivos, debido a la acogida que tuvo, que a propósito fue la misma que ha tenido a nivel mundial.

Pregunta No. 3 Para usted Cloud Computing es:

Cuadro # 3

Orden	Alternativas	f	%
1	Un paradigma que permite ofrecer servicios de computación a través de Internet.	10	71%
3	El uso de equipos de telecomunicaciones y computadoras para la transmisión, el procesamiento y el almacenamiento de datos.	3	21%
4	No sé.	1	7%
TOTAL		14	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí. **Elaborada por:** Verónica E. Muentes Molina

Gráfico # 3

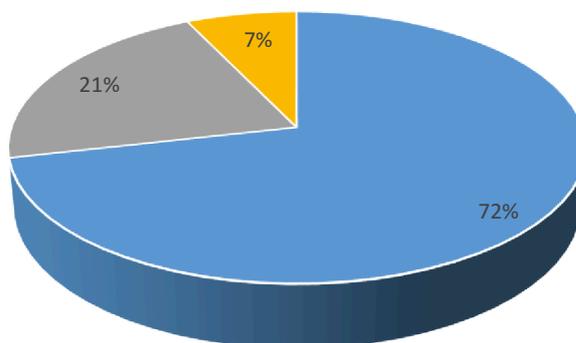


Ilustración 23 Encuesta 3 - Pregunta3

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, un 72% conocen sobre Cloud Computing y un 28% desconoce; como nos indica el gráfico podemos observar que la computación en la nube es un término muy conocido actualmente, y todo ello debido al antes y después que marcó a partir del internet.

Pregunta No. 4 ¿Ha escuchado o ha usado alguna de las siguientes plataformas de servicios de Cloud Computing? Encierre en un círculo.

Cuadro # 4

Orden	Alternativas	f	%
1	Microsoft Azure.	11	44%
2	Google Cloud Platforms.	7	28%
3	IBM SmartCloud.	3	12%
4	AWS.	1	4%
5	No he escuchado, ni usado.	3	12%
TOTAL		25	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 4

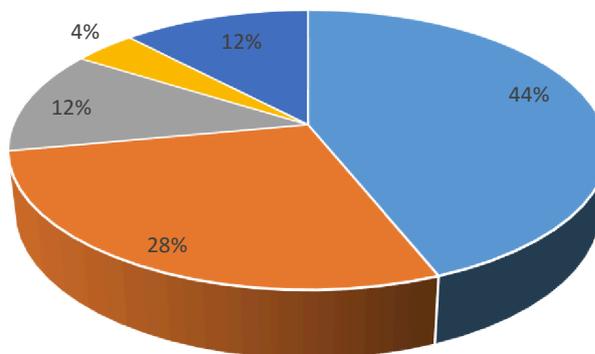


Ilustración 24 Encuesta 3 - Pregunta4

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, el 44% ha escuchado o ha usado Microsoft Azure, otro 44% han usado o escuchado sobre otras plataformas cloud y un 12% no ha escuchado, ni usado. Si podemos ver el gráfico, Azure es la más conocida si se toma en cuenta que el 44% es representado sólo por MA. Entonces es una ventaja.

Pregunta No. 5 Señale las ramas de Inteligencia Artificial que usted ha escuchado o conoce.

Cuadro # 5

Orden	Alternativas	f	%
1	Redes Neuronales.	13	27%
2	Lógica Difusa.	7	15%
3	Algoritmos Genéticos.	11	23%
4	Robótica.	12	25%
5	Machine Learning.	3	6%
6	Otras.	1	2%
7	No he escuchado, ni conozco.	1	2%
TOTAL		48	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 5

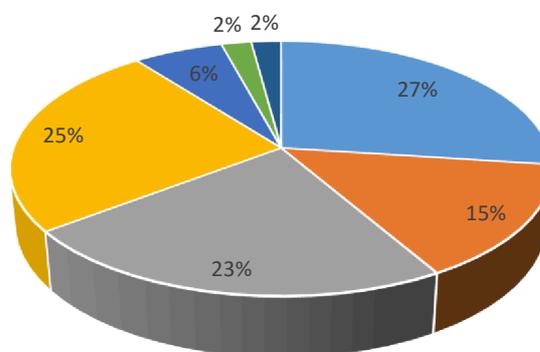


Ilustración 25 Encuesta 3 - Pregunta5

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, el 90% ha escuchado o conoce sobre las ramas más sobresalientes de la Inteligencia Artificial, excepcionando al Machine Learning que sólo obtuvo el 6%; lo que denota que existe un bajo porcentaje en cuanto a conocimientos sobre el Aprendizaje automático, una de las ramas más importantes de la IA.

Pregunta No. 6 Según su criterio encierre una de las opciones, según lo que piensa es lo más importante para tener una predicción precisa en Machine Learning.

Cuadro # 6

Orden	Alternativas	f	%
1	Datos lo más normalizados posibles.	1	7%
2	Algoritmo eficiente.	6	43%
3	Gran cantidad de datos pasados.	3	21%
4	No sé	4	29%
TOTAL		14	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 6

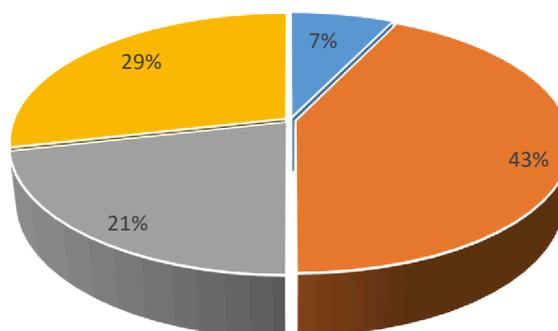


Ilustración 26 Encuesta 3 - Pregunta 6

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, el 43% determinó que un algoritmo eficiente es lo más importante en el ML, un 7% determinó que lo era una buena normalización y un 21% acertó al considerar la base de datos pasados como el determinante. Según estos resultados, se puede deducir que sólo un 7% conoce más a fondo sobre ML, y el otro 79% tiene un conocimiento bajo sobre esta nueva tecnología.

Pregunta No. 7 Sabe usted, ¿Cuál de las siguientes opciones son tipos del Machine Learning?

Cuadro # 7

Orden	Alternativas	f	%
1	Aprendizaje Supervisado.	3	19%
2	Parallel.	1	6%
3	Aprendizaje por posición.	2	13%
4	Transducción.	0	0%
5	No sé	10	63%
TOTAL		16	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico # 7

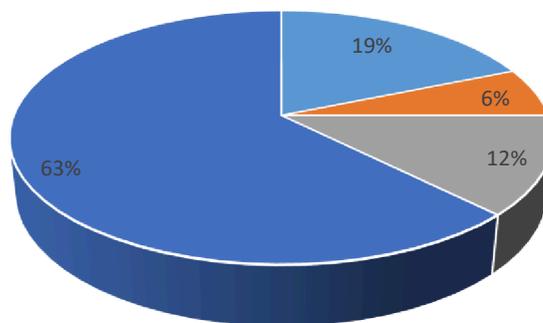


Ilustración 27 Encuesta 3 - Pregunta 7

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, el 82% desconoce sobre tipos de ML dejando abierta una posibilidad por ampliar conocimientos.

Pregunta No. 8 Si ha hecho experimentos o aplicado el Machine Learning. ¿Bajo qué entorno para análisis de conocimiento lo ha hecho? Marque con x.

Cuadro # 8

Orden	Alternativas	f	%
1	Machine Learning en Microsoft Azure.		0%
2	Machine Learning en Weka.	1	7%
3	Machine Learning en Python.		0%
4	En otro entorno.		0%
5	No he aplicado el machine Learning aún.	13	93%
TOTAL		14	100%

Fuente: Docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

Elaborada por: Verónica E. Muentes Molina

Gráfico #8

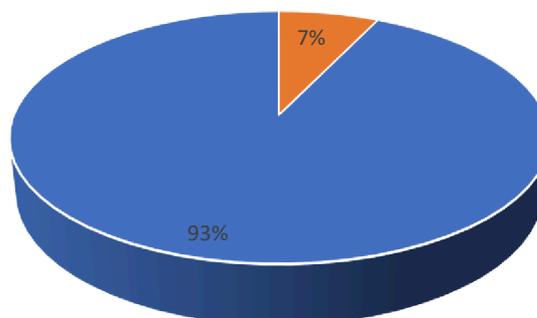


Ilustración 28 Encuesta 3 - Pregunta 8

Análisis: Según encuestas realizadas a docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, El 93% de los de los docentes de las áreas técnicas de la Facultad de Ingeniería de Sistemas de la ULEAM, no han aplicado aún el Machine Learning. Es común el resultado debido a que la aplicación de ésta tiene una reciente acogida.

2.9.2. Informe final del análisis de los resultados.

En base a la información recolectada por la técnica de investigación de encuestas, se concluyó que los estudiantes de últimos semestres de la carrera de Ingeniería de Sistemas de la ULEAM ubicada en la ciudad de Manta después de su análisis y recopilación tuvieron los siguientes resultados:

En la primera encuesta, se puso observar que tienen conocimientos sobre el Cloud Computing, dando una base para poder incursionarse en plataformas de servicios, y tienen además breves conocimientos sobre la IA que les va a ayudar a tener claro a que pueden orientar el ML. Se puede decir que actualmente no cuentan con un conocimiento básico de Machine Learning, pero muchos de ellos desarrollan aplicaciones web o móviles que podrían transformar si conocieran más sobre cómo usarla.

En la segunda encuesta sobre actitudes, se pudo analizar que la mayoría de estudiantes tienen planes a futuro en el sector, algunos ya ponen en práctica sus conocimientos. El ambiente se presta a la investigación, sólo necesitan exigirse más. El valor del Machine Learning puede dar valor a sus emprendimientos. Necesitan abrir el mundo de opciones en su cabeza con nuevo conocimiento, y de esta forma tener más experiencia.

En la tercera encuesta sobre el conocimiento de los docentes sobre la IA, el Cloud Computing y el Machine Learning se puede observar que existe una base al igual que en los estudiantes sobre la cual poder fundamentar el Machine Learning en Azure; sin embargo es notorio el desconocimiento sobre ML específicamente, algo que como mencionaba anteriormente no es para menos, debido a la reciente aceptación que ésta teniendo en la actualidad. Aunque ya existe hace muchos años, es ahora que con la Minería de datos, el Big Data y la Inteligencia de negocios que ha cogido fuerza, y está cambiando las cosas.

Respecto a lo anteriormente expuesto, el capítulo II ratifica la necesidad de una guía sobre Machine Learning aplicada en Azure, plataforma catalogada como la estrella de Microsoft, con el fin de proveer una base de confianza y conocimiento para la actualización de conocimientos y dar valor a las aplicaciones que se desarrollan por los estudiantes. ML está dando mucho de qué hablar y es capaz de despertar la curiosidad de cualquiera. Aprender a hacer experimentos con machine Learning es una arma capaz de potenciar nuevos emprendedores, y proyectos de investigación.

CAPITULO III: DISEÑO DE LA PROPUESTA

3.1. INTRODUCCIÓN

Este capítulo fue generado en base al análisis de los resultados de las encuestas, con la intención de dar a conocer a los estudiantes universitarios, las funcionalidades y servicios que una herramienta como Microsoft Azure puede ofrecer, gracias a la tecnología y la Computación en la Nube, proponiendo una guía escrita y virtual estructurada de forma clara y concisa, que comprenda dos secciones; una, la base teórica para comprender Machine Learning y la otra una base práctica en cuanto a su uso por medio de la plataforma Azure de Microsoft.

Fusión de los segmentos y resultado del estudio de la investigación, que sirvió para dar a conocer la adopción de soluciones en cuanto a predicción de información y aprendizaje automático se refiere, con el fin de la revelación de beneficios que se pueden obtener con el consumo de servicios por medio de plataformas cloud, facilidades que posee y el valor que agregan al software.

3.2. TITULO DE LA PROPUESTA

Guía sobre el uso de Machine Learning basado en “Regresión Lineal”, orientado a la plataforma de servicios en la nube Microsoft Azure.

3.3. DESCRIPCIÓN DE LA PROPUESTA

En base a la problemática de la investigación, se procedió a plantear la solución más acertada que permita alcanzar los objetivos planteados. La creación de una guía básica para aprender sobre Machine Learning como rama

de la Inteligencia Artificial, y como servicio de Microsoft Azure, dividida en dos segmentos, teoría y práctica.

La primera etapa estuvo fundamentada en el análisis de información recaudado en la investigación y en los diferentes usos que le pueden dar los estudiantes de la carrera.

La segunda etapa, trató sobre experimentos hechos por terceros y propios, explicaciones claras para aplicar el ML y los costos por el servicio de Aprendizaje Automático.

Proporcionando por medio de éstos dos segmentos una base de investigación para entender el funcionamiento de Azure, las diferentes aplicaciones que se le puede dar y como poner en práctica los experimentos.

3.4. JUSTIFICACIÓN DE LA PROPUESTA

El Machine Learning es una rama de la Inteligencia Artificial con muchos años desde su descubrimiento, que recientemente ha tomado fuerza gracias a la Inteligencia de Negocios, el Data Mining y el Big Data. Este tema se trata mucho en la actualidad a nivel tecnológico, hay cualquier cantidad de cursos en internet porque se está promoviendo su aprendizaje.

Los estudiantes, al igual que los docentes de la Facultad de Ciencias Informáticas de la ULEAM de Manta, tienen un conocimiento bajo en lo que respecta a este tema, perdiéndose la transformación que ML puede dar a sus soluciones, así como el uso que se le puede dar independientemente. Dado el caso, es importante para el desarrollo estudiantil de los alumnos, conocer sobre nuevas tecnologías y herramientas que faciliten el hacer de la carrera, y por qué no, que por medio de éstas además se les pueda dar vida a los proyectos informáticos con los que se cuenta para darles un valor extra.

3.5. OBJETIVOS

3.5.1 Objetivo General

- Aumentar los niveles de conocimientos sobre Machine Learning en el Cloud Computing mediante regresión Lineal aplicado en Microsoft Azure.

3.5.2 Objetivos Específicos

- Socializar y diseñar una guía para el uso de Machine Learning basado en “Regresión Lineal”, por medio de la plataforma de servicios en la nube Microsoft Azure a los líderes estudiantiles y docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí.

3.6 PROPUESTA

Introducción.

Antes de que existiera la inteligencia de negocios, la minería de datos y el Big Data, ya se manejaba el Machine Learning. Era muy complicada su aplicación y quienes la llevaban a cabo eran equipos de físicos-matemáticos, ingenieros especializados en Data Science, ingenieros especializados en inteligencia artificial y expertos en la solución lógica.

Una empresa con un departamento especial para hacer Machine Learning representaba mucho dinero, conocimiento y tiempo, debido a que podía llegar a tomar años dar con una hipótesis; incluyendo que manejar información en grandes volúmenes es bastante complejo. Con el paso del tiempo apareció el Cloud Computing y con ella nuevas maneras para el tratamiento y procesamiento de datos, por lo que en la actualidad ya no es tan complejo, ni costoso como antes.

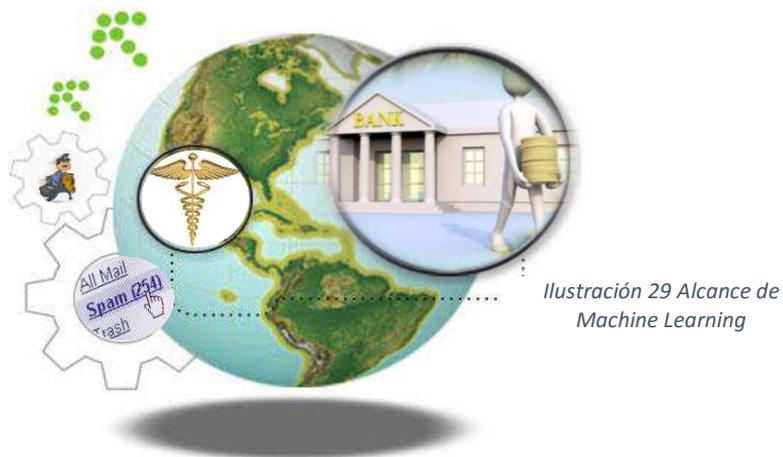


Ilustración 29 Alcance de Machine Learning

Computación en la Nube

Es un modelo de tecnología que permite el acceso a los servicios de computación (Software, Plataformas, Infraestructura), por medio del Internet.

En los últimos años se ha convertido en un triunfo para el mundo como lo conocíamos, debido a su escalabilidad, movilidad y coste. En el pasado, cumplir con los requerimientos y la implementación exigía un amplio presupuesto; sin embargo con el también llamado cloud computing, la cultura de consumir servicios ha permitido que esté al alcance de todos, ayudando al desarrollo de la educación, de las empresas y de la tecnología.



Ilustración 30 Estructura de la computación en la nube

Plataformas Cloud más populares.

2006



Ilustración 31 Amazon Web Services

2009



Ilustración 32 IBM Smart Cloud

2010



Ilustración 33 Microsoft Azure

2012



Ilustración 34 Google Cloud Platform

Microsoft Azure

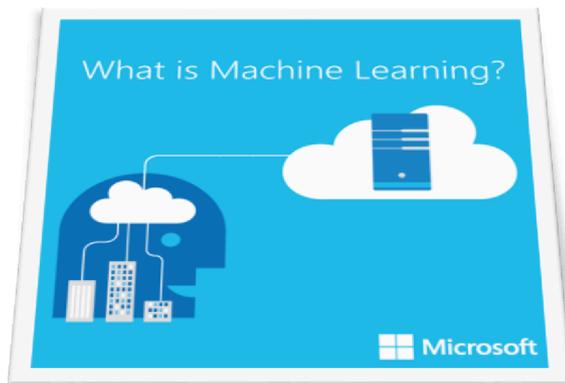


Ilustración 35 ¿Qué es Machine Learning?

Azure es la plataforma de informática en la nube de Microsoft, es de las más grandes plataformas de servicios existentes en la nube, cuenta con una variedad de servicios integrados, entre ellos análisis, bases de datos, móviles, redes, almacenamiento y web.

Es una plataforma intuitiva e interoperable, lo que facilita su uso a nivel general, sin mencionar que facilita

un sin número de tareas como la compilación, la administración de aplicaciones, la integridad con otros sistemas operativos, lenguajes en cuanto a software. Y en cuanto a hardware ha sido reconocido por Gartner en ser el mejor en Infraestructura como Servicio (IASS), por su facilidad para acoplarse a otras infraestructuras, permitiendo contar con nubes híbridas para las empresas que aún no se fían de tener todo en la nube, así como las ISO's que cumple.

La ejecución de Azure se basa en una red mundial de datacenters administrados por Microsoft en 19 regiones. Estamos hablando de más países y regiones que Amazon Web Services y Google Cloud juntos.

Machine Learning

A simples rasgos el Machine Learning trata de construir sistemas de cómputo que puedan adaptarse, aprender de las experiencias que le otorguen los datos históricos, y de las experiencias que esta misma pueda ir teniendo.

Se debe tener claros los conceptos de percepción y aprendizaje. El primero no es más que la manera en cómo se organiza, identifica e interpreta la

información sensorial, con el fin de representarla y entenderla. El segundo es el hallazgo de una hipótesis que dado otros datos, luego se va haciendo más precisa.

Las hipótesis se miden con una función de costo y entre más mínimo sea el error cuadrático, será más preciso.

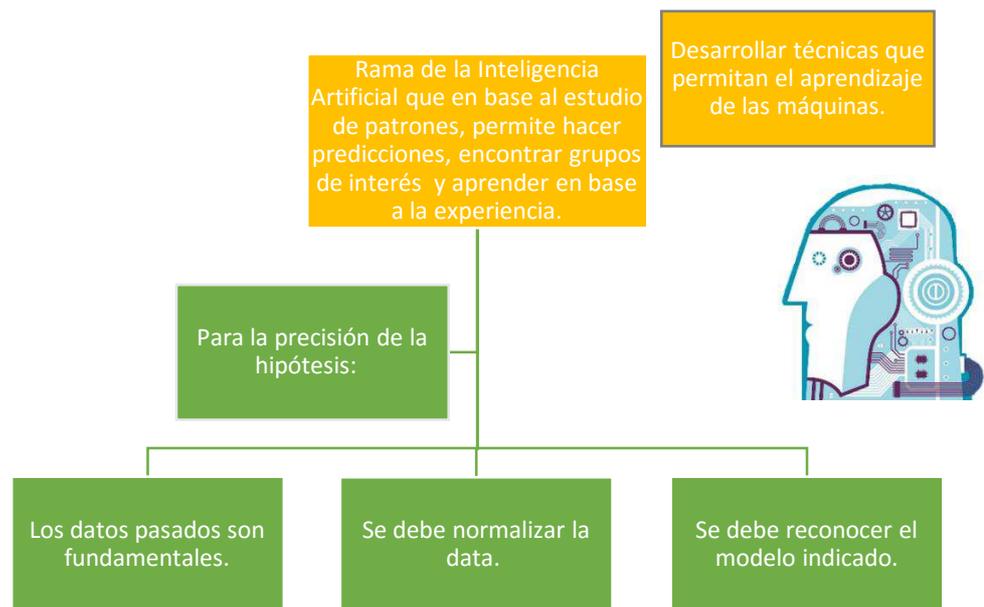


Ilustración 36 Características de Machine Learning

¿Aprendizaje supervisado? ó ¿Aprendizaje no supervisado?

Al escuchar estos dos términos muchos han de pensar que el aprendizaje supervisado consiste en que el ser humano debe estar pendiente todo el tiempo de que aprenda, programándolo y apilando botones y que el no supervisado es uno automático. Lo cierto es que no funcionan así, depende totalmente de lo que se quiera predecir para poder escoger que tipo se aplica.

El aprendizaje supervisado...

Se utiliza cuando la predicción requiere específicamente como resultado una variable establecida o un valor en especial.

Este aprendizaje está dividido en dos subdivisiones:

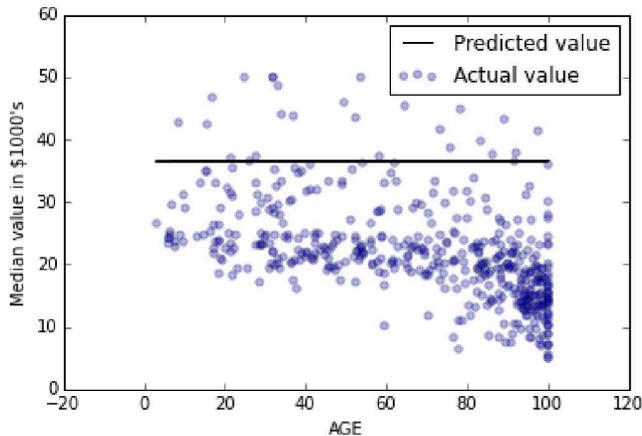


Ilustración 37 Modelo de predicción de valor

Regresión: Para predecir valores continuos, por ejemplo cuánto ganaré el próximo año.

Clasificación: Para obtener valores discretos. Por ejemplo personas que tienen o no diabetes.

El aprendizaje no supervisado...

Se utiliza cuando se quieren buscar patrones que determinen nuevos grupos o estructuras como salidas, y para ser más concreta, cuando se sabe poco o nada sobre los valores a predecir. Este aprendizaje se divide en:

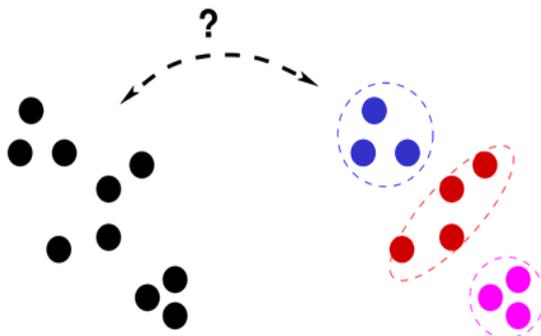


Ilustración 38 Asociación

Clúster: Partiendo de una colección de datos, se busca agruparlos por distintas variables como sexo, nombre, ingreso, etc.

Asociación: De un conjunto de datos, busca los que tengan cierta asociación con algo, por ejemplo dolor de cabeza y de estómago son síntomas de x enfermedad.

¿Cómo funciona la Regresión Lineal?

La regresión Lineal es una herramienta estadística que estudia fenómenos en diversos campos de la ciencia.

En aprendizaje automático, forma parte de los métodos de aprendizaje, en su caso del aprendizaje supervisado, que busca variables de entrada que se ajusten a una hipótesis para partir de ésta y dar un pronóstico. Caracterizado por buscar relaciones entre variables y marcar una pendiente a la que determina hipótesis.

La hipótesis será pronosticada con su función:

$$h_{\theta}(x) = \theta_0 + \theta_1 x$$

Lo que es equivalente a que digamos $y = mx + b$

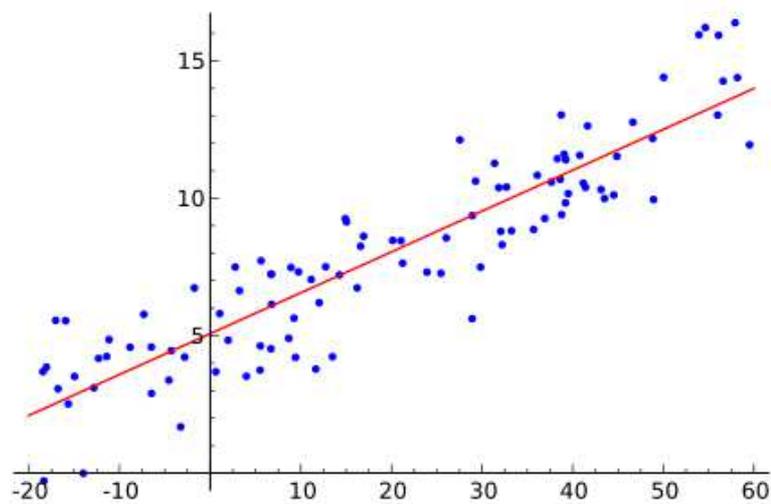


Ilustración 39 Regresión lineal

Entonces lo que hace es asignar valores a theta cero y theta uno, con los cuáles se pueda obtener un valor de y para x.

Pero, ¿Cómo conocer la precisión de la hipótesis?

Se la puede conocer por medio de una función de costo equivalente al método de la suma de cuadrados. Este toma el promedio de todos los resultados

aplicando la función hipótesis y los compara con los reales obtenidos, sumando los errores positivos y negativos al cuadrado.

$$J(\theta_0, \theta_1) = \frac{1}{2m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})^2$$

Los errores negativos son todos los puntos que se ubican por debajo de la pendiente y los positivos los que son igual o están por encima de la línea. Obtendremos un valor numérico que entre más cercano a cero indicará si la hipótesis es la mejor.

¿Cómo poder escoger los valores adecuados para theta o y theta 1?

Para poder minimizar la función de costo u obtener el valor más cercano a cero se utiliza un algoritmo llamado gradiente descendente, el cual en cada iteración se desplaza en la dirección que lo acerca al punto más bajo, para ello lo multiplica por la tangente.

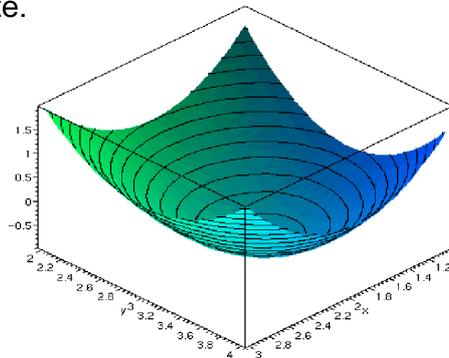


Ilustración 40 gradiente descendente

Se Repite hasta que convergen:

$$\theta_0 := \theta_0 - \alpha \frac{1}{m} \sum_{i=1}^m (h_{\theta}(x^{(i)}) - y^{(i)})$$

$$\theta_1 := \theta_1 - \alpha \frac{1}{m} \sum_{i=1}^m ((h_{\theta}(x^{(i)}) - y^{(i)})x^{(i)})$$

Pero, ¿Qué hace Microsoft Azure con Machine Learning?



Ilustración 41 Machine Learning

Microsoft Azure Machine Learning combina el poder de un aprendizaje de máquina completo con los beneficios de la nube.

Automatiza todo lo el proceso de cálculos que viste anteriormente, facilitando la normalización de data y modelando tus estrategias con eficiencia y eficacia por medio de la nube y con un sistema totalmente administrado.

Lo más importante del proceso es la **data**, más que los modelos de entrenamiento y demás factores importantes; pero no cualquier data, deben estar limpia y con la mínima cantidad de campos en blanco.

¿Por qué la data es lo más importante?

- Ingeniería de factores.
- Entrenamiento de conjuntos.
- Ciencia de Datos. (Data Science)

Ventajas y Desventajas de Microsoft Azure Machine Learning.

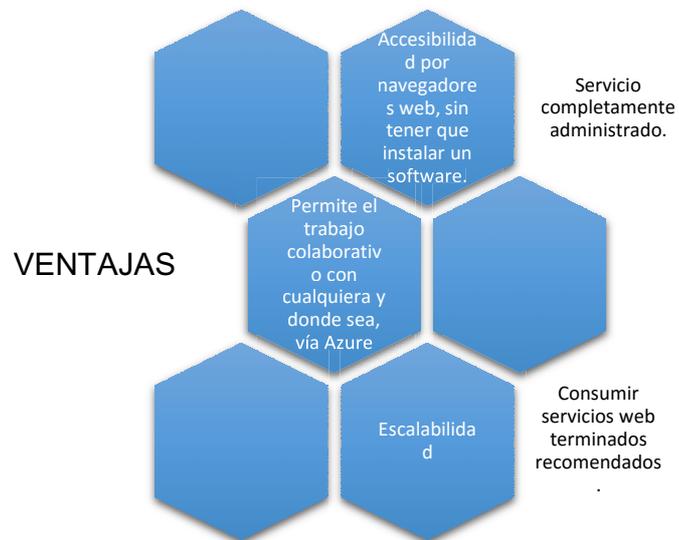


Ilustración 42 Ventajas Machine Learning



Ilustración 43 Desventajas Machine Learning

Machine Learning y Minería de Datos.

Machine Learning y la Minería de Datos son consideradas como primas hermanas por ser tecnologías similares basada en patrones, y aunque una descubre patrones y la otra los busca. Lo cierto es que sin la minería de datos no sería tan fácil para el Machine Learning hacer predicciones tan precisas. Porque si el machine Learning no conoce el patrón a buscar, deberá asumir.

Por ejemplo el Big Data, éste puede analizar grandes volúmenes de datos (Gigabytes, Terabytes, Petabytes y Exabytes) descritos como estructurados, no estructurados y semiestructurados, pero todo esto gracias al Data Mining.

Entonces pueden ver que se usa en conjunto con otras tecnologías, y hace más de lo que parece. **Es el medio para el fin.**

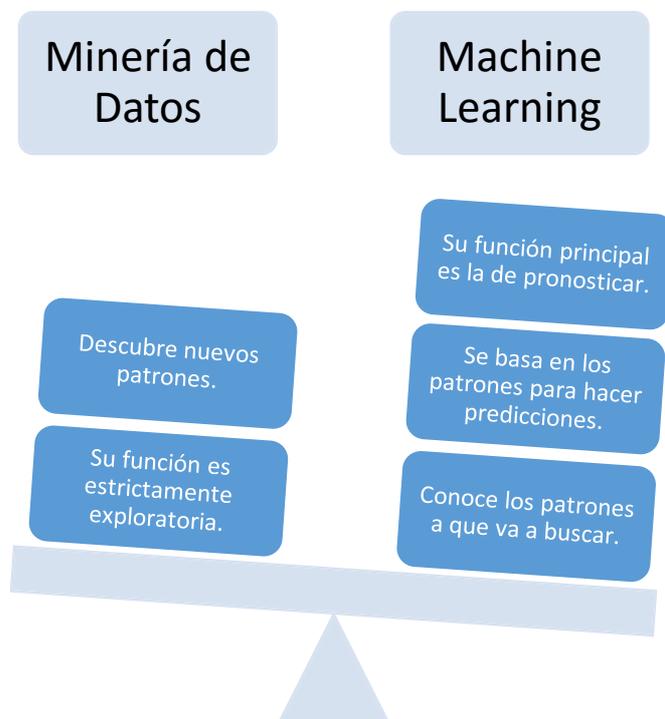


Ilustración 44 Minería de datos vs Machine Learning

Características para determinar si tu solución debe ser Machine Learning.



Ilustración 45 Servidores de computadoras

Debes tener en cuenta que no vas a buscar una solución en Machine Learning para cualquier requerimiento. Por ello a continuación te presento características para que determines si se aplica o no a tus necesidades.

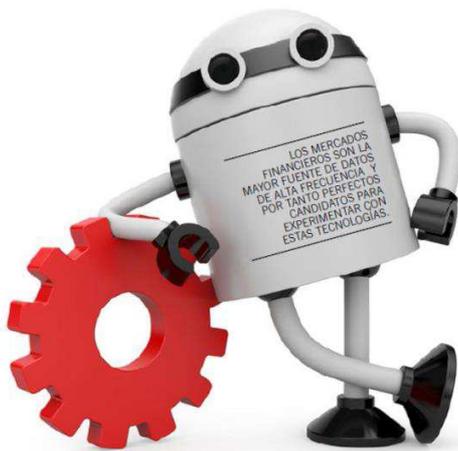


Ilustración 46 Necesidades Machine Learning

- ✓ Se debe tener datos históricos disponibles.
- ✓ Su núcleo son predicciones automáticas.
- ✓ Sin demasiadas reglas de negocios que gobiernen la experiencia.
- ✓ Sistemas corrientes dedicados a las predicciones de un valor, o grupo de valores.

Pasos para construir una solución Machine Learning en Microsoft Azure.



Ilustración 47 Pasos para construir una solución Machine Learning

Estudio Machine Learning.

Es una herramienta fácil e intuitiva que por medio de sus funciones de arrastrar y colocar, permite la construcción de soluciones de análisis predictivos, alimentándose con datos; combina la ciencia de datos, el análisis predictivo, y los recursos en la nube.



Ciencia de Datos

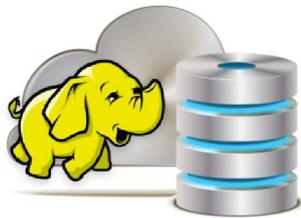
Blobs

Tables

Hadoop (HDInsight)

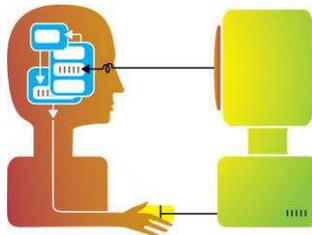
Bases de datos

Relacionales (Azure SQL
DB)



Modelo Predictivo

Entorno de Desarrollo
de Machine Learning.



Web API operacional
en minutos.

Se puede obtener el
modelo através de un
servicio web(web
service) listo para ser
llamado.



Ilustración 48 Estudio Machine Learning

Introducción al Estudio ML.

Al ingresar al estudio, podrás observar las siguientes opciones del menú principal.

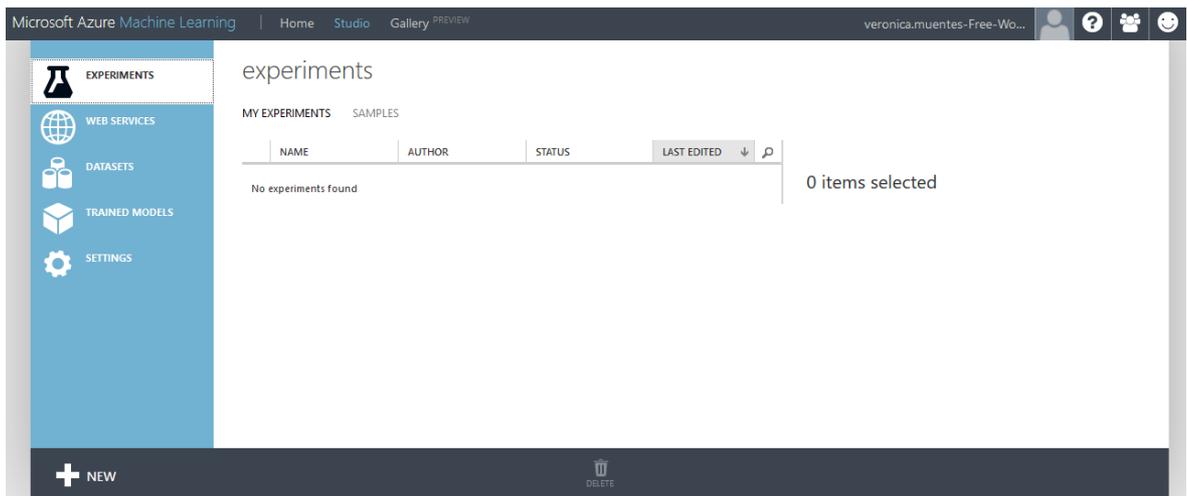


Ilustración 49 ML Estudio

Ubicada en la parte lateral izquierda:

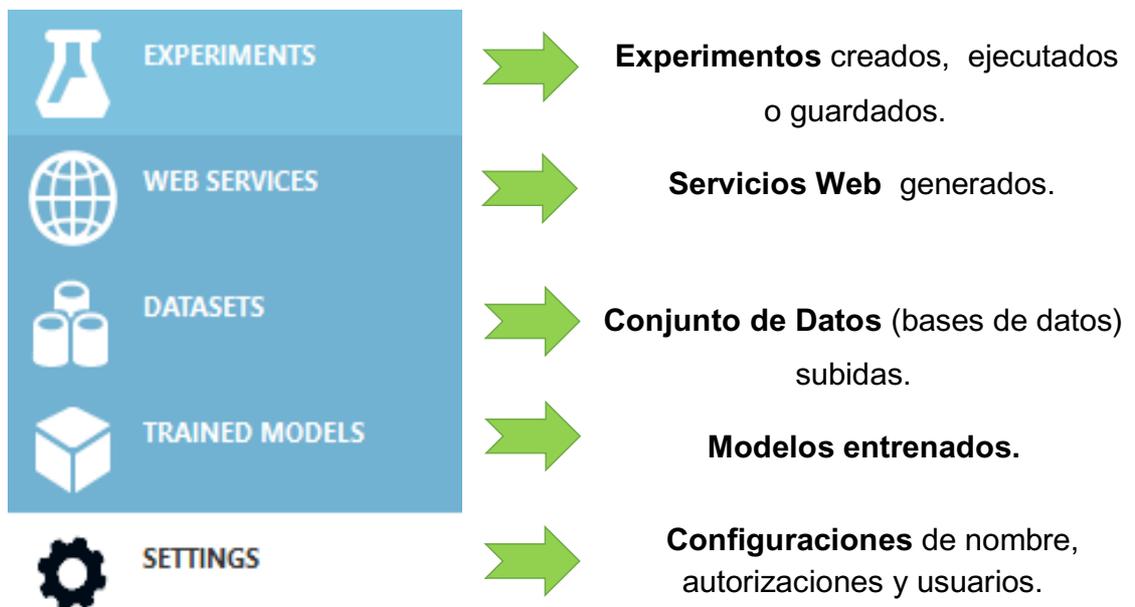


Ilustración 50 Opciones ML

Ubicada en la parte inferior:

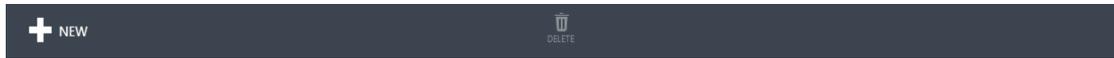


Ilustración 51 Nuevo ML

Sirve para crear un nuevo experimento con sólo dar un click en **+NEW**

Al crear un nuevo experimento, te encontrarás con una lista de trabajo donde se pueden observar conjuntos de datos y componentes con sus propiedades dispuestos para la construcción del modelo.

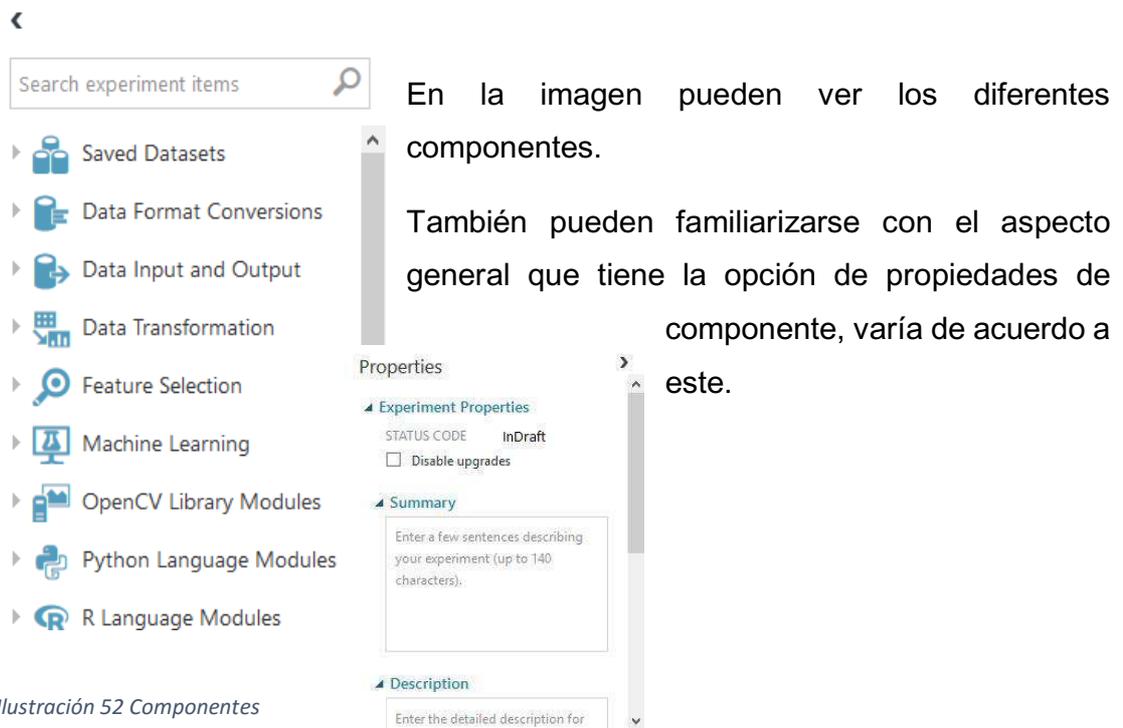


Ilustración 52 Componentes

El lienzo de experimentos cambia una vez que estamos dentro de un experimento, se encuentran más opciones, como guardar, correr, preparar servicios web, entre otros.



Ilustración 53 Opciones 2

Práctica en el Studio ML.

Se creará un modelo de regresión lineal para predecir el precio de un automóvil en base a las diferentes variables de entrada, como marca y otras especificaciones técnicas.

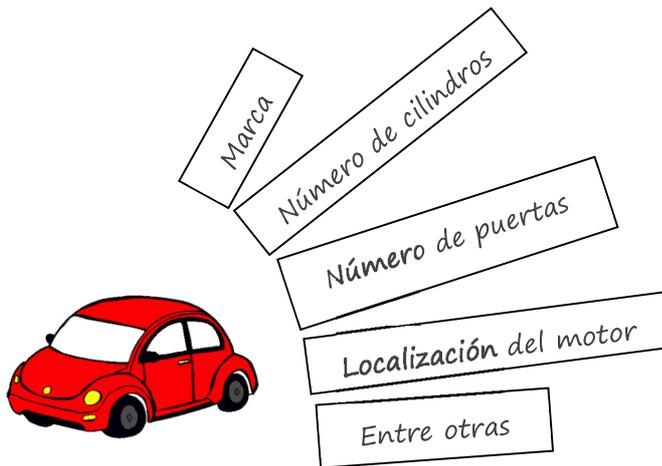


Ilustración 54 Características auto

Paso 1. Ingresar al Studio ML.

Opción 1. Ingresar por el medio de tu cuenta Microsoft Azure.

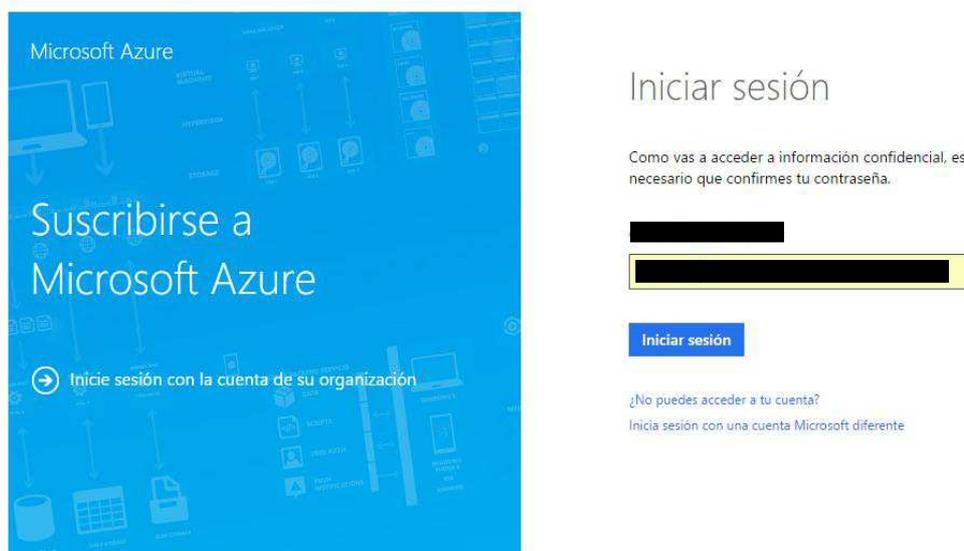


Ilustración 55 Iniciar sesión

Después de loguearte, ingresas al portal y al encontrar estos servicios, escogerás **Aprendizaje Automático** y **Abrir en Studio**.

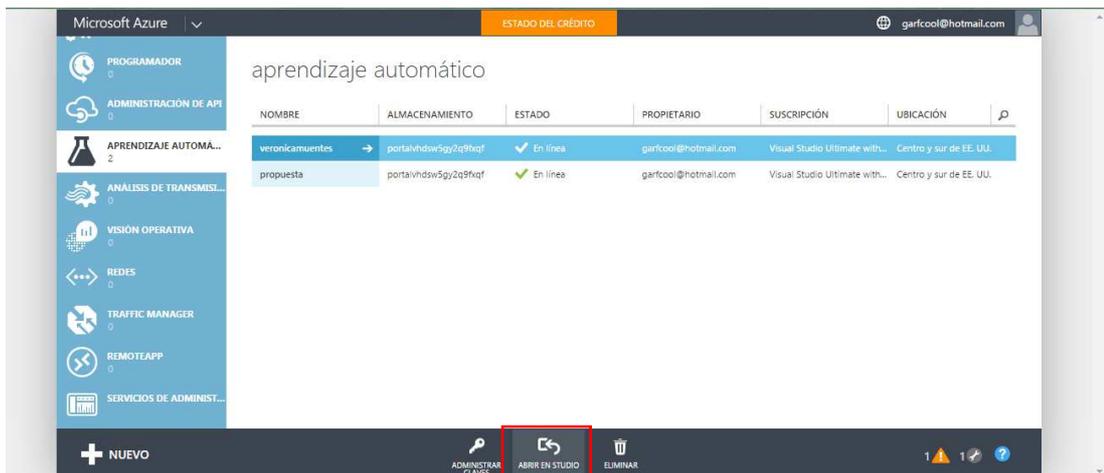


Ilustración 56 Abrir en Studio

Esto puede ser diferente para cada usuario. Si cuentas con suscripción, eres un Microsoft Student Partner, te piensas acoger o probar un mes, ingresas así; se debe tener en cuenta que al probar por un mes se deben ingresar los datos de una tarjeta de crédito, pero sin cobro alguno mientras se cumpla el tiempo gratis.

Pero, si deseas probar sin ningún compromiso de tarjetas de crédito, puedes ingresar por medio de la opción 2.

Opción 2

Ingresas en el siguiente link:

<https://azure.microsoft.com/es-es/documentation/articles/machine-learning-create-experiment/>

Encontrarás algo como esto, entonces debes dar click en la opción de comience ahora y listo.

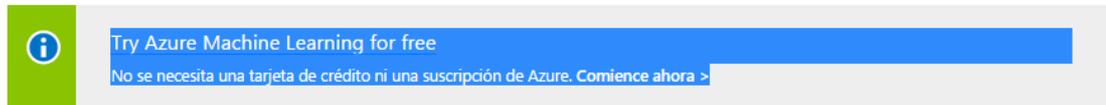


Ilustración 57 Probar Machine Learning

Paso 3. De cualquier forma vas a ingresar al Studio ML y tendrás algo como lo que ves a continuación.

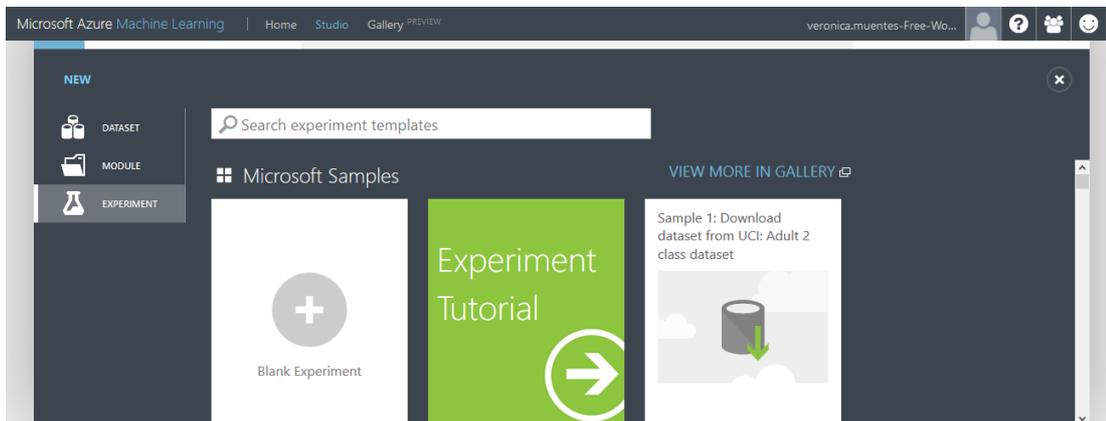


Ilustración 58 ML Studio

Tendrás tres opciones

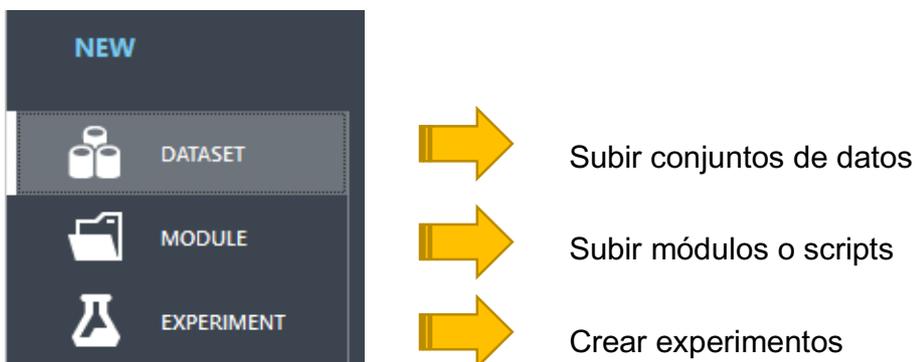


Ilustración 59 Nuevo ML

La **primera opción** es para subir archivos locales de bases de datos adicionales.

Pueden observar los formatos que acepta Studio ML para la alimentación de data, desde CVS files, texto plano, hasta archivos ZIP; podemos observar su alto nivel de compatibilidad.

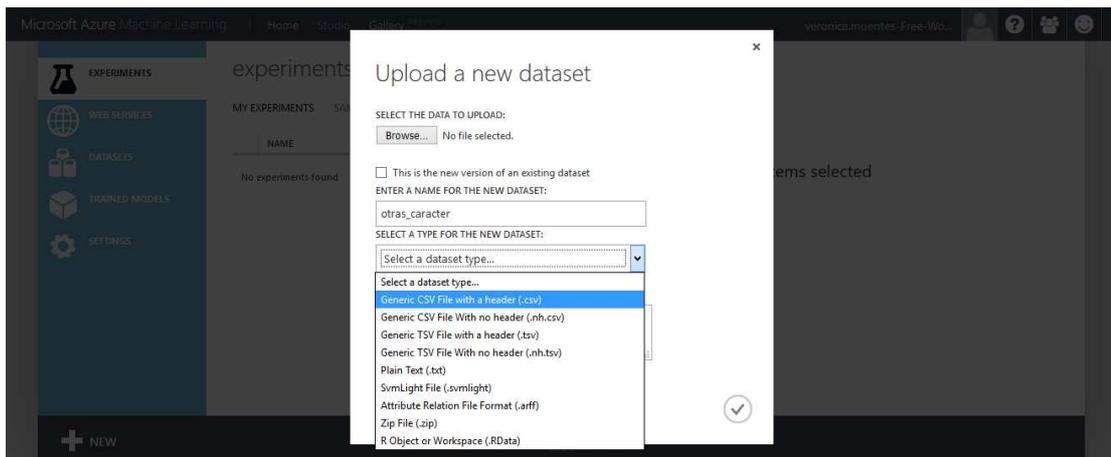


Ilustración 60 Cargar Dataset

La **segunda opción** es para subir módulos adicionales escritos en R o Python, lenguajes soportados por Studio ML.

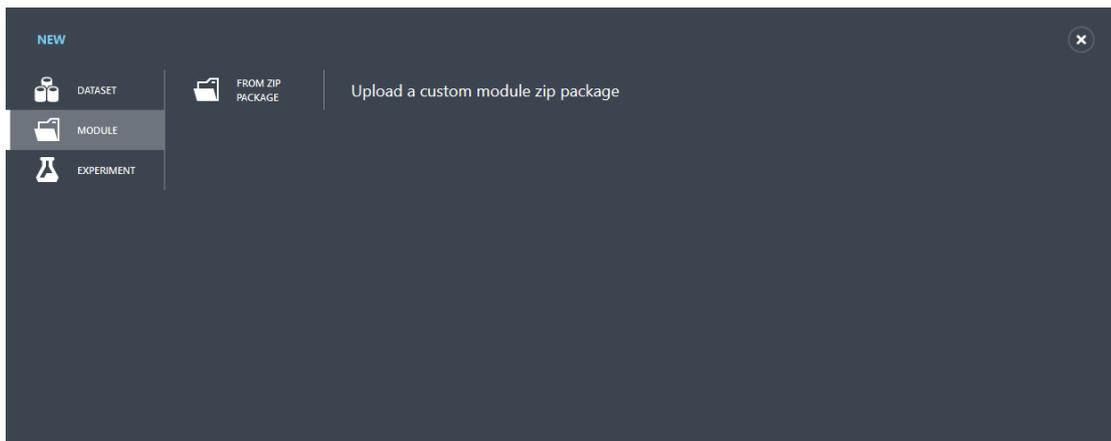


Ilustración 61 Cargar Módulo

La **tercera opción** es para crear un nuevo experimento, entonces en esta sección, vamos a darle click en la opción <Blank Experiment>

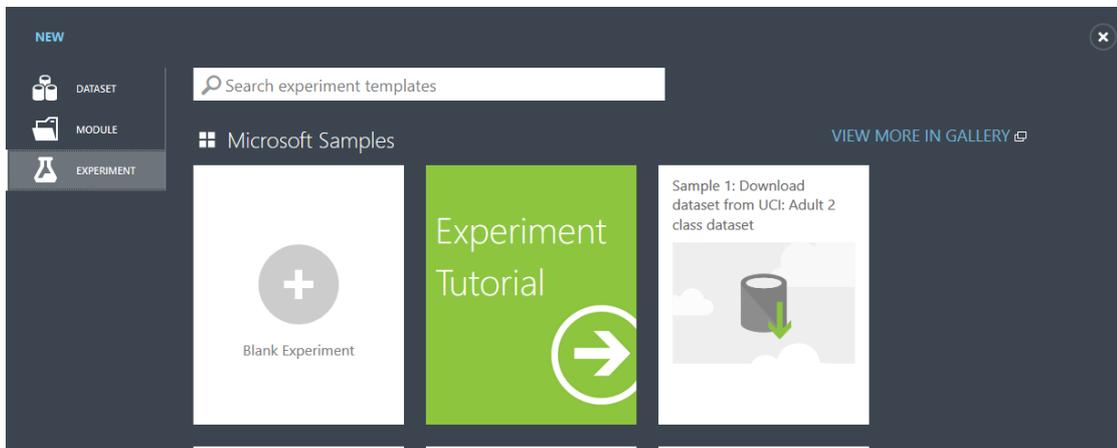


Ilustración 62 Experimento

Obtendremos algo como esto:

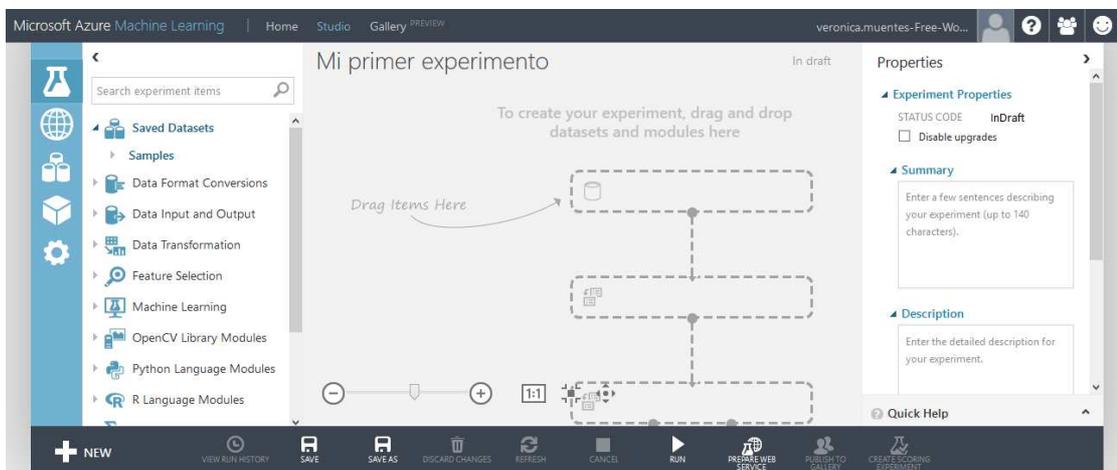


Ilustración 63 Nuevo experimento

Ahora que sabemos las funciones del menú principal, sus componentes y módulos, crear un nuevo experimento y entrar en nuestro Studio ML. Vamos a empezar a crear el modelo.

Paso 4. Procedemos a realizar los pasos para construir una solución ML. Subir la data.

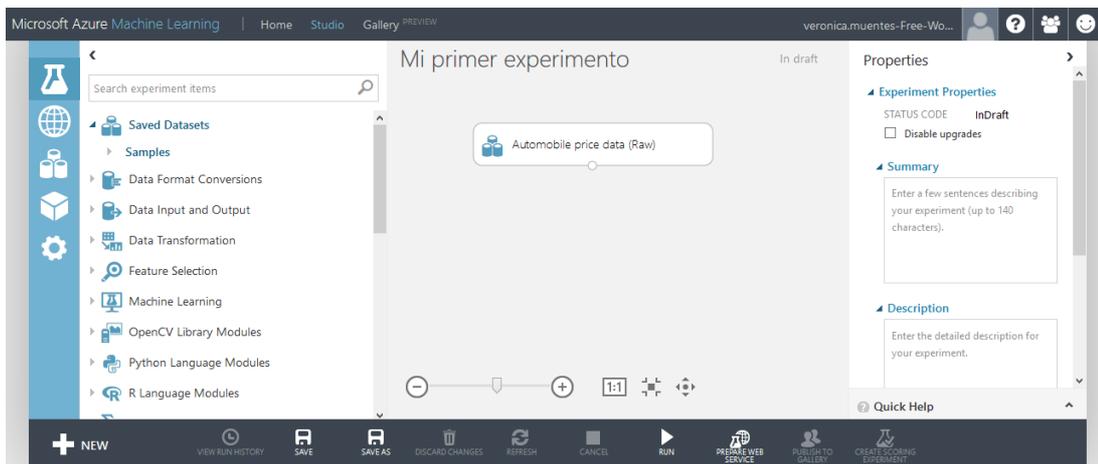


Ilustración 64 Primer experimento

Si acercamos el puntero del mouse al componente, se mostrará esta opción.



Ilustración 65 Selección de DataSet

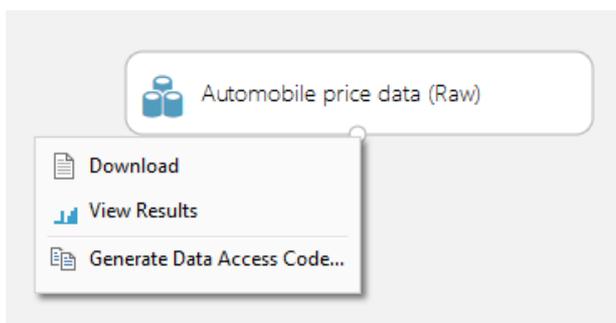


Ilustración 66 Opciones DataSet

Si le damos click, tendremos opción a descargar la data, ver los resultados desde el Studio y generar código de acceso a la data.

Paso 6. En proyectar columnas, se pueden eliminar columnas que son irrelevantes en la predicción, para mantener lo más limpia la información.

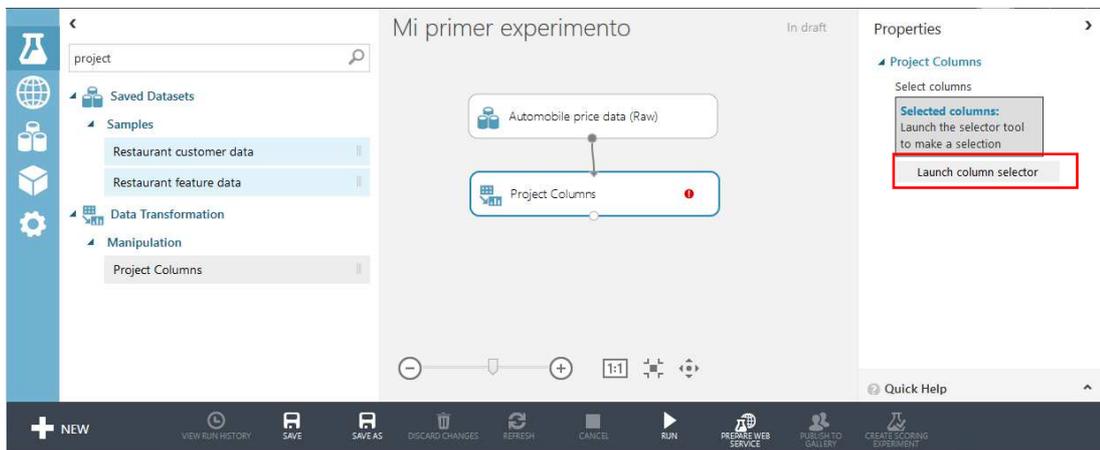


Ilustración 67 Proyectar columnas

Se seleccionarán las columnas dando click en el <Launch column selector>

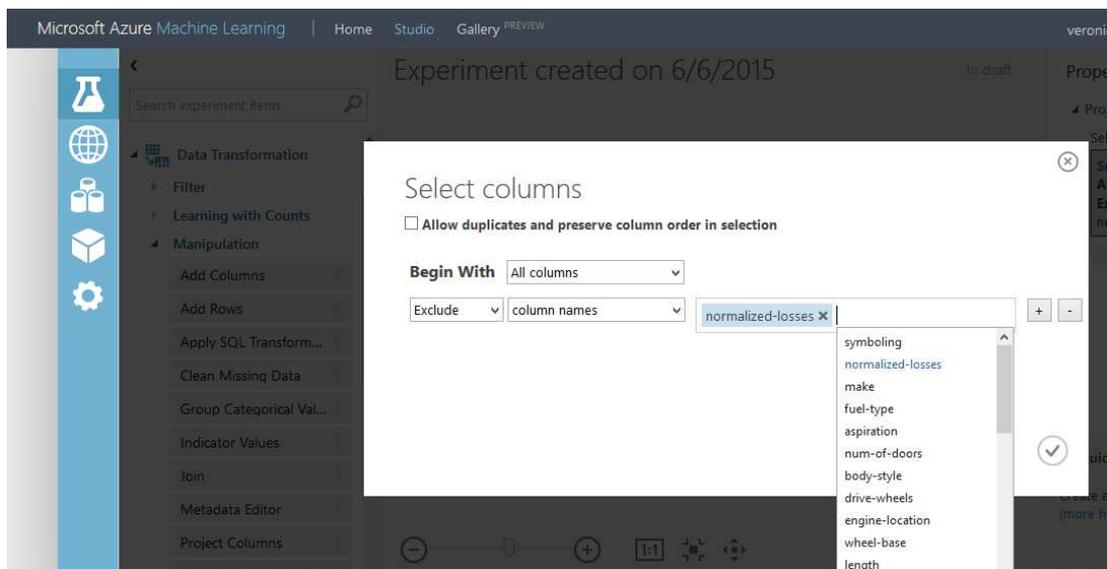


Ilustración 68 Excluir campos



IMPORTANTE:

Múltiples sitios web proporcionan accesos a los grandes datos de forma barata y con etiquetado. Pero la calidad también importa en el ML.

Paso 7. Eliminar filas con valores null.

Por defecto al ubicar la opción <Missing Values Scrubber> en el diagrama, para los valores nulos, se sustituirán por defecto con un valor que nosotros le podemos programar como se ve en el gráfico.

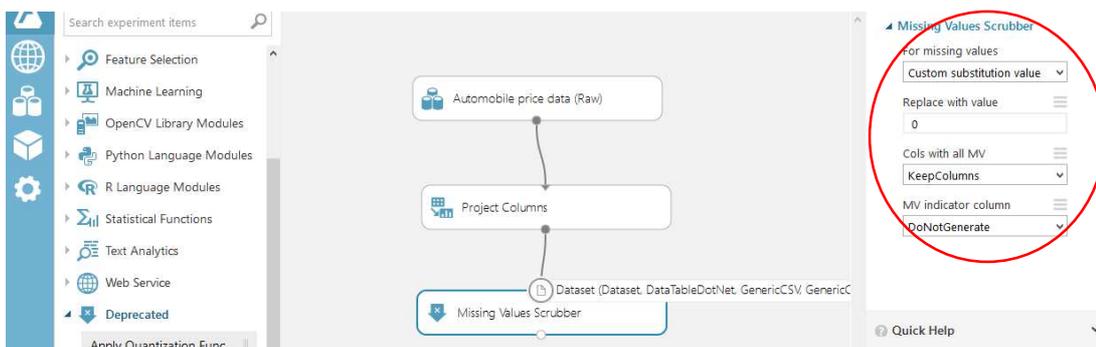


Ilustración 69 Sustitución de valores nulos

Sin embargo cambiaremos el valor por defecto y los sustituiremos por <Remove entire row> y sus valores por defecto.

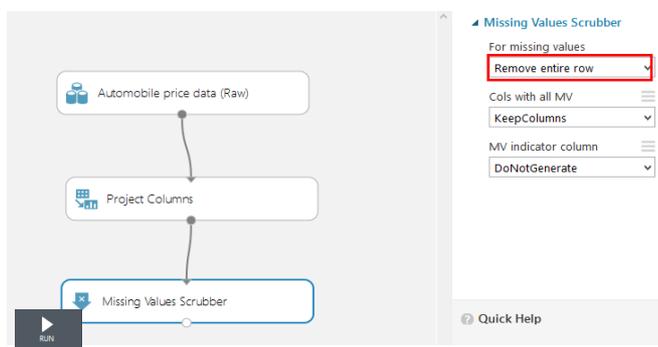


Ilustración 70 Valores nulos

Esto permite que se elimine toda la fila que tiene valores null, para mantener lo más limpia posible nuestra data.

Luego vamos a ejecutar el experimento, dando click a la opción de Run que se puede

ver en el lienzo de experimentos, para poder limpiar los datos.

Una vez ejecutado podremos visualizar ahora dos flechas verdes que sirven de verificación, esto sólo indica que la limpieza se ha completado correctamente.

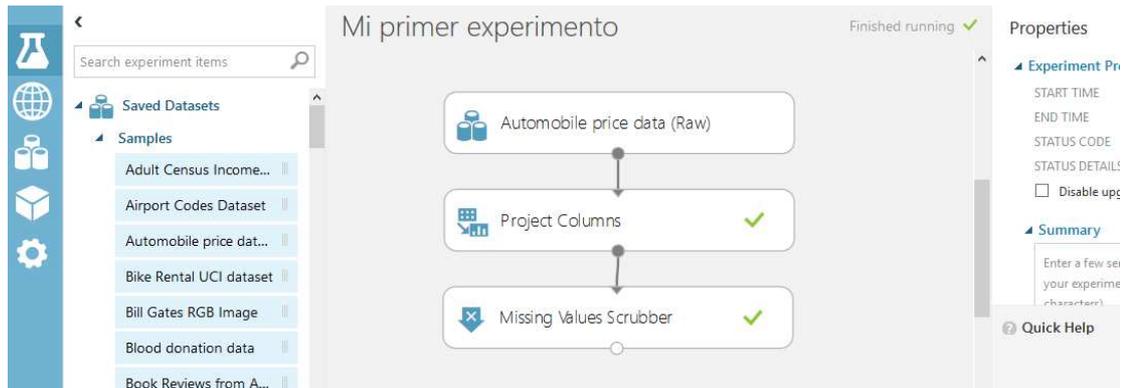


Ilustración 71 Ejecución de experimento

Paso 8. Partiendo de que cada fila representa a un carro y cada columna las diferentes características de este carro, en este paso es necesario definir las mejores características para la predicción agregando un <Project Columns> en el que se ingresen los atributos que creemos son las más relevantes para predecir el precio y listo.

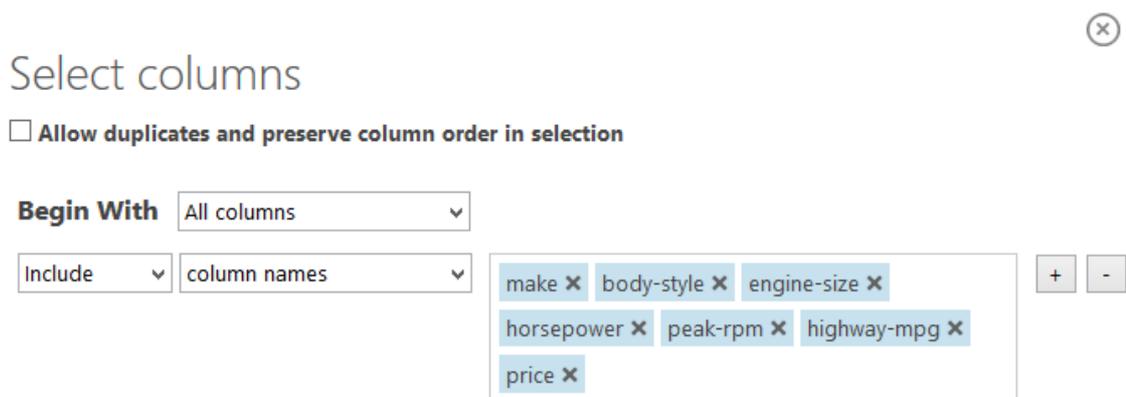


Ilustración 72 Seleccionar columnas

Paso 9. Realizaremos una división entre los datos que serán utilizados para el entrenamiento del modelo y los datos para la predicción. Es importante ubicar una fracción de filas, que será utilizada para el entrenamiento.

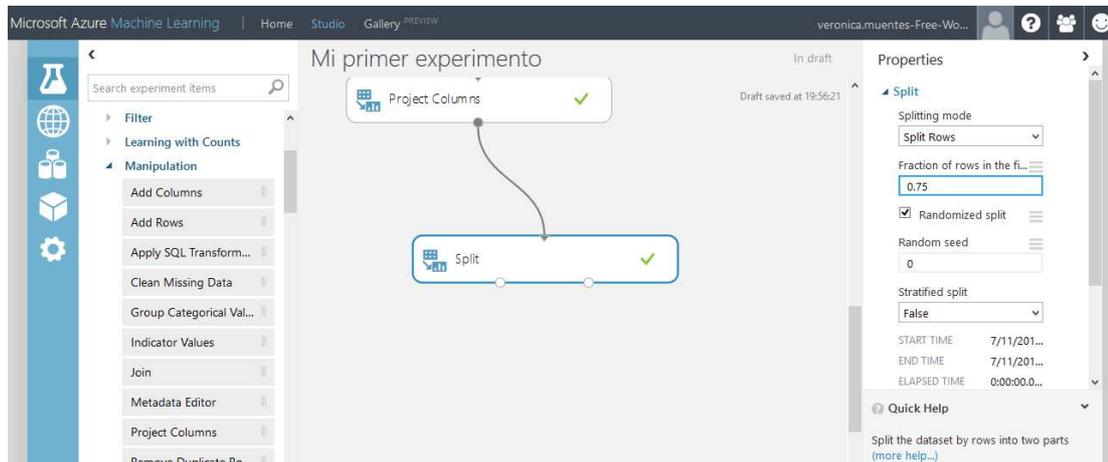


Ilustración 73 División

Se recomienda que para mejores resultados se deje un 75% de datos para el entrenamiento.



Paso 10. Ahora que tenemos los datos listos, podemos empezar a entrenar el modelo. Entonces seleccionamos el componente <Train Model> y el <Score Model>, los arrastramos al modelo.

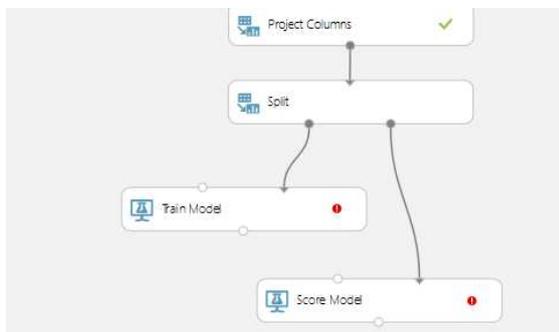


Ilustración 74 Modelo de entrenamiento

El <Split> enviará un 75% para entrenarse, en este caso al <Train Model>, y el 25% para ser probado y medido en el <Score Model>

Paso 11. Una vez se hayan dividido, deberemos asignar el modelo que va a usarse en el entrenamiento. En este caso “Linear Regression”, y en propiedades seleccionar el método de gradiente desc.

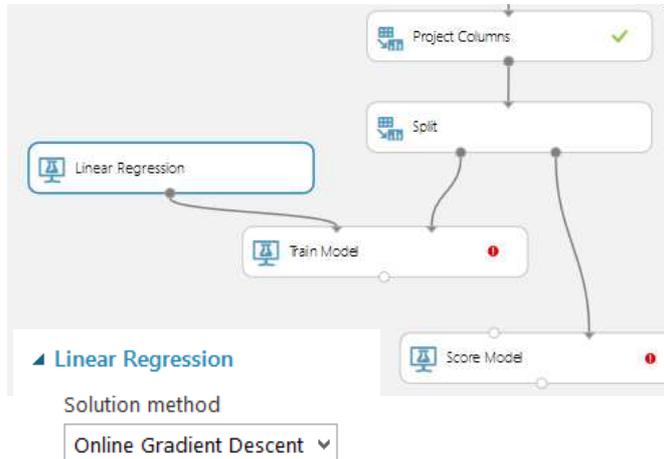


Ilustración 75 Modelo de puntaje

Select a single column

Include column names price X

Podemos observar una alerta en el <Train Model> requiriendo el valor a entrenar, en este caso se lo agrega. (precio)



Ilustración 76 Campo de entrenamiento

Una vez terminado el modelo, ya podemos ver la predicción.

Para poder visualizarla **ejecutamos el modelo.**

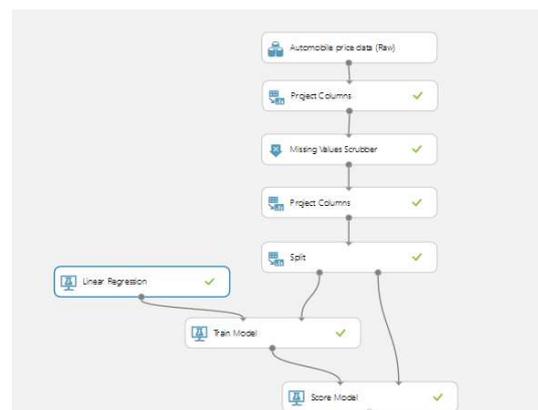
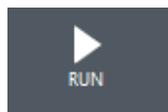


Ilustración 77 Correr modelo

Esto toma unos minutos, y luego la verificación con flechas verdes confirmará que ha sido efectuado



Paso 12. Podemos comparar resultados visualizando la data. Click en el círculo y visualizar datos.

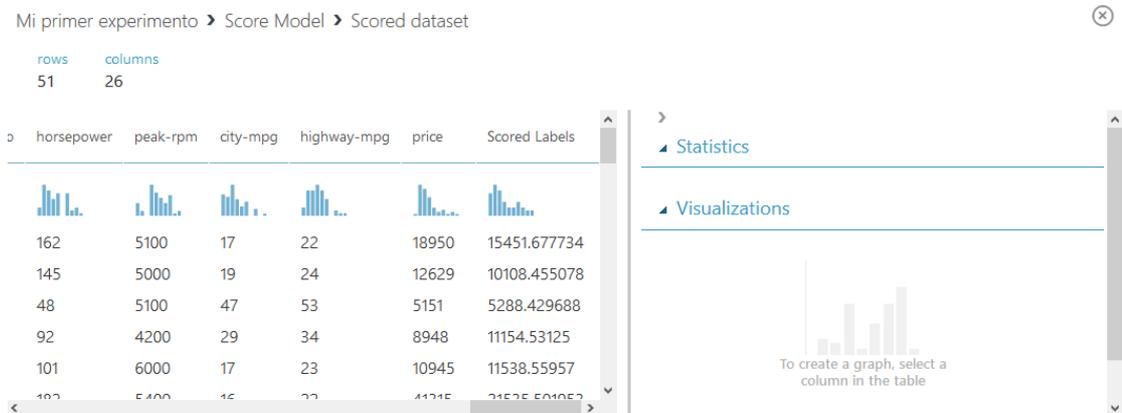


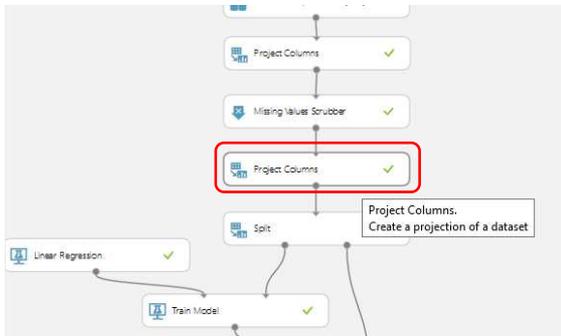
Ilustración 78 Visualización de datos

i **IMPORTANTE:**

Es fundamental que la persona que realice el experimento lo haga junto a un especialista en el caso a estudiar, para obtener mejores resultados.

El <price> de la imagen es el precio que tenía realmente el carro (\$18950,00), y el modelo predecía que cuesta \$15451,68. Está muy bien el resultado, pero el modelo puede llegar a ser más preciso si se lo entrena mucho más. No esperemos que a la primera iteración nos salgan los valores exactos, se requiere de entrenamiento y de conocimiento por parte nuestra en la materia del estudio.

¿Cómo lo entrenamos?



Esto se hace en el segundo Project Columns, donde incluimos las columnas creímos eran las más importantes, podemos sacar una e ingresar otra, y volver a ejecutar, no hace falta eliminar, basta con irse a propiedades y modificar las columnas

o características.

Existen un sin número de modelos para entrenamiento, se los escoge de acuerdo a la necesidad, como les pude explicar al principio.

Paso 13. Evaluaremos de manera precisa la calidad de los resultados obtenidos, agregando el componente <Evaluate Model> y volviendo a correr el experimento.

Visualizamos los resultados y podremos obtener los márgenes de error. Oficialmente tenemos el modelo terminado.

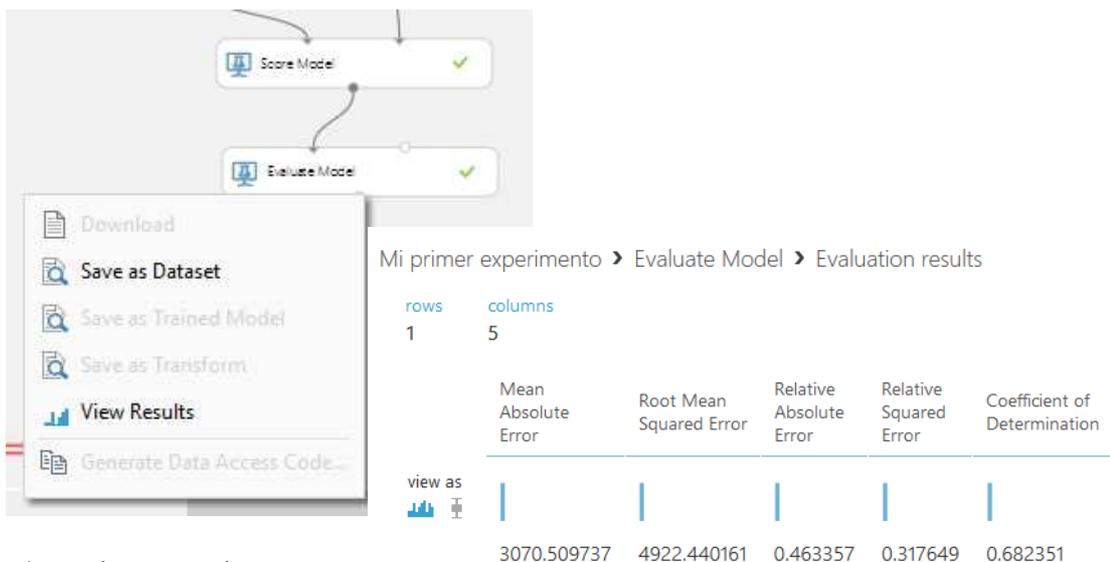


Ilustración 79 Ejecución ML

Adicionales...

Publicar como Webservice

Si queremos publicar el modelo como servicio web para consumirlo desde alguna de nuestras aplicaciones, en el lienzo sólo escogemos la opción de <Set up Web Service>, y luego <Predictive Web Service>

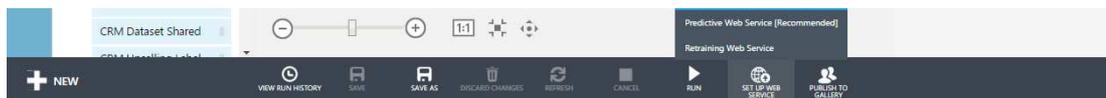


Ilustración 80 Web Services

Automáticamente se crea un entorno de <Predictive experiment> a lado de nuestro <Training experiment> creado anteriormente, Podremos ver que automáticamente se han ubicado los componentes de <Web Service input> y <Web Service output>, sólo vamos a ejecutar.



Ilustración 81 Respuesta Web Services

Encontraremos algunas opciones en el lienzo de abajo. Clickearemos la opción de <DEPLOY WEB SERVICE>, para desplegar el servicio web y se abrirá automáticamente el tablero de nuestro servicio web.



Ilustración 82 Web Services de experimento

Encontraremos entre lo más importante para consumir el servicio web:

La API Key:



Ilustración 83 Api Key

Su uso y funcionamiento:

-Dando click en REQUEST/RESPONSE

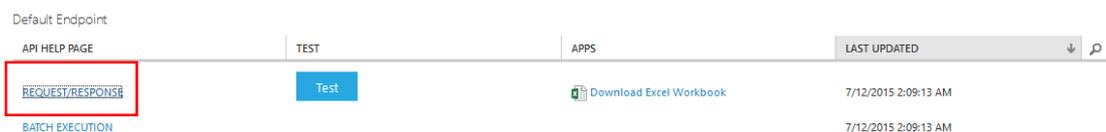


Ilustración 84 Respuesta de solicitud

-Se abrirá una página completa sobre información de la API.

Request

Method	Request URI	HTTP Version
POST	https://ussouthcentral.services.azureml.net/workspaces/b2030d70821c4daa2392f354aa3f476/services/67726b14de8b4a2ca4c17f770a9fc6b9/execute?api-version=2.0&details=true	HTTP/1.1

Note: You may omit the **details** parameter from the query string. This would cause **ColumnTypes** to be omitted from the output

Request Headers

Request Header	Description
Authorization:Bearer abc123	Required. Pass the API Key here. Obtain this key from the publisher of the API.
Content-Length	Required. The length of the content body.
Content-Type:application/json	Required if the request body is sent in JSON format.

Ilustración 85 Solicitud de Web Services

Con esta información ya podemos hacer una prueba de que el servicio web funciona descargando un proxy de depuración web como **Fiddler Web debugger** que es gratuito o uno equivalente para probar las entradas y salidas, o directamente probarlo desde un sitio web.

¿Qué puedes hacer con ML?



Ilustración 86 Búsqueda

Entre sus usos más usuales encontramos:

- Correo postal.
- Resultados Médicos.
- Publicidad Dirigida.
- Análisis de rotación.
- Motores de Búsqueda.



Ilustración 87 Análisis



Ilustración 88
Buscadores

- Pronóstico del Tiempo.
- Análisis de Redes Sociales.
- Mantenimiento predictivo.
- Detección de intrusos en la red.
- Indicaciones de Tránsito vía GPS.
- Optimización de aplicaciones web.
- Exploración de Recursos Naturales.
- Modelo de Tendencias de Compras.
- Detección de Fraudes en los bancos.
- Recomendaciones en línea de productos.
- Asistentes personales en móviles, laptops como Cortana.



Empresas que usan ML en Azure.

Seattle, E.E.U.U



Ilustración 89 MAX 451

Se dedica a la predicción de productos que los clientes querrán en el futuro y a crear campañas de marketing más eficientes y eficaces.

California, E.E.U.U



Ilustración 90 OSIsoft

Se dedica a la detección de desperfectos en tiempo real y en diagnósticos de variaciones en la energía de salida.

Alemania, Essen y Duisburgo



Ilustración 91 ThyssenKrupp

Se dedica a la fundición, bienes de inversión y servicios. Permiten el monitoreo en tiempo real de sus ascensores, entre otros controles.

Pensilvania, E.E.U.U



Ilustración 92 Theorem

Se dedica a ofrecer servicios integrales de investigación y desarrollo clínicos.

Costos por uso del Studio Machine Learning.

En Microsoft Azure sólo se paga por lo que se consume, el costo siempre será proporcional al tiempo de uso, y en el caso de tener en producción servicios web, el costo será por número de transacciones

Ilustración 93 Costos Azure



\$7.44 por usuario



\$0.75 por hora



\$0.38 por 1000 transacciones de API de producción. (mínimo)



\$0.05 por el segundo por predicción, de proceso de API de producción.

CAPITULO IV: EVALUACIÓN DE RESULTADOS

4.1 INTRODUCCIÓN

La tecnología evoluciona día a día manteniéndonos siempre al pendiente de las nuevas innovaciones que se desarrollan, permitiéndonos mejorar los procesos diarios que utilizamos, mientras este crecimiento sea exponencial las posibilidades en cuanto a precio y uso aumentarán. En la actualidad hemos llegado al punto de poder predecir valores, comportamientos, resultados y acciones, y siguen perfeccionándose los métodos.

Para poder lograr esto, es necesario el componente de actualización e investigación de los individuos que conforman una sociedad, por medio de documentos como esta guía, que permiten un acceso a una fuente de investigación especializada en una determinada tecnología.

4.2 SEGUIMIENTO Y MONITOREO DE RESULTADOS

- **Seguimiento.**

Por medio de la guía se dio a conocer a estudiantes líderes de comunidades estudiantiles, al presidente estudiantil y a profesores de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí, el impacto y la utilidad del Machine Learning y su facilidad de experimentación gracias a Microsoft Azure. También se intentó dar conciencia del impacto que tiene en nuestras vidas diarias, y porque estudiarla.

Entre más conocimientos tengamos, mayor será el desarrollo y con ello, el alcance de mayores comodidades a bajo costo en nuestro diario vivir. Antes los medios y recursos eran limitados, pero con la computación en la nube y las

plataformas cloud, todos deberíamos interesarnos por conocer un poco más y aplicarlo en nuestras aplicaciones, implementaciones o negocios para tomar mejores decisiones.

Se efectuó también una encuesta corta para medir cuanto habían aprendido y se obtuvieron resultados bastante buenos; sin embargo esta guía no busca sólo aumentar el conocimiento en los encuestados, o en alumnos de niveles superiores, por ello está dirigida al público en general, para cualquier persona que quiera saber más sobre esta tecnología, y espero o los

- **Monitoreo.**

La Guía sobre el uso de Machine Learning basado en “Regresión Lineal”, orientado a la plataforma de servicios en la nube Microsoft Azure, muestra las dos caras del Machine Learning y como con el avance tecnológico, ya no tiene el mismo grado de complejidad que tenía en el pasado.

Según noticias, periódicos y revistas desde el año 2014 esta tecnología empezó a ofrecerla Microsoft, Google, entre otros, y la mayoría de empresas sólo observaban, la querían pero la creían costosa, desconociendo la situación actual. Estamos 2015 y se puede leer, escuchar que un sin número de empresas están empezando a implementarla, se ha reestructurado totalmente la forma en la que antes se veía al aprendizaje automático, y se ha vuelto un servicio al alcance de todos.

Por otra parte, era significativo el desconocimiento de Machine Learning en el medio, no sólo por parte de las empresas, sino también por parte de los profesionales y estudiantes, por lo que con la guía se pensó en contrarrestar el desconocimiento que existía sobre este tema. Y los propósitos fueron:

- **Propósito:**

Aumentar los niveles de conocimiento sobre Machine Learning basado en “Regresión Lineal”, orientado a la plataforma de servicios en la nube Microsoft Azure tomando en cuenta los primeros resultados:

- **Resultados:**

Se logró conseguir un nivel de aumento en el conocimiento de los estudiantes bastante aceptable, teniendo en cuenta que no conocían en su mayoría sobre esta tecnología; según resultados anteriores y resultados obtenidos luego de la socialización de la guía, se puede tener por seguro que hay menos ignorancia respecto al tema, y que saldrán más profesionales experimentados en el uso de Machine Learning, expandiendo así el conocimiento a todo el que esté dispuesto a darle unos tiempos libres. Qué sería de la tecnología si cada día no hay más investigadores y profesionales desarrollándola.

Los avances y resultados de las metas logradas, así como los cuadros estadísticos levantados durante la ejecución del proyecto, servirán como antecedentes del conocimiento disgregado.

Orden	Componentes	Indicador situación anterior	Indicador situación actual
1	Conocimiento sobre Cloud Computing	87%	95%
2	Conocimiento sobre la plataforma cloud Microsoft Azure	50%	95%
3	Conocimiento sobre la división del aprendizaje automático.	1%	50%
4	Conocimiento sobre Machine Learning.	36%	90%

Tabla 1 Indicador de situación

- Gráfica.



Ilustración 94 Resultados antes del estudio

Resultados obtenidos de la Encuesta 1, realizada con el fin de determinar un nivel real de los conocimientos sobre Machine Learning en Azure.

Resultados obtenidos de la encuesta de seguimiento con respecto a la socialización de la guía.



Ilustración 95 Resultados después del estudio

CONCLUSIONES

- En base al análisis hecho al estudiante promedio, y debido al escaso material en español sobre este tipo de tecnologías, se concluye que la guía presentada para el aprendizaje de Machine Learning en Azure es de gran valoración en los estudiantes y el público en general,
- La regresión lineal, entre otros modelos usados en el Machine Learning, son bastante complejos. Entre más variables, aumenta el nivel de dificultad del modelo; sin embargo en la actualidad no se necesita ser un erudito en física y matemáticas para aplicarlo. Entender el funcionamiento básico sobre “Linear Regression” es difícil si no encuentras información clara, por eso he plasmado en la guía una explicación al entendimiento de todos.
- La teoría siempre va a proporcionar una base y a veces detalles que sólo los aprendes leyéndola, pero la práctica siempre termina por ayudarte a entenderla en su totalidad. Entonces la práctica siempre será lo mejor para terminar de comprender un concepto, todo va de la mano.
- El servicio de Machine Learning en Azure es como una renta, su uso es proporcional a su costo. Entre más horas lo usas, pagas más. Entonces su valor está bastante bien, porque siempre se lo encuentra ajustado a la medida; sin embargo es necesario señalar que Microsoft tiene programas universitarios que fomentan la investigación, lo que permite que se pueda solicitar como institución el servicio por un año gratis, siempre que se fundamente el proyecto que requiere esta tecnología.

RECOMENDACIONES

- Es necesario darle el tiempo que requiere la lectura y práctica de la guía, para obtener los resultados que se desean tanto práctica como teóricamente, además es recomendable que si existe mayor interés por seguir haciendo experimentos y llevarlos a producción, se debe leer más al respecto, tomando en cuenta que el Machine Learning es una rama amplia y que existen varios modelos según su problemática que se aplican en el Studio ML.
- Es necesario conocer el funcionamiento básico del modelo que se va a aplicar en Machine Learning, para que no existan sorpresas y más que todo porque sólo así aprendes a valorar la tecnología y la facilidad que te ofrecen otras herramientas hoy en día, como Microsoft Azure y su Estudio ML.
- Los experimentos de Machine Learning necesitan ser llevados por personas expertas en el campo del problema, sean profesionales o no, porque lo más importante es la limpieza del ruido de la data, y ésta sólo se logra exitosamente cuando se conoce el problema a cabalidad. No se pueden esperar resultados precisos, si no se tiene total conocimiento; por eso al menos que sólo sean experimentos de prueba no se necesitan a gran rasgo, pero si se habla de llevar a producción el modelo (WebServices de Azure), es mejor contratar a alguien que lo sea, para no desperdiciar el dinero.
- Tener clara la situación problemática y publicar el modelo sólo cuando se está seguro de que es el más preciso, para evitar la pérdida de tiempo y transacciones, que en Azure, equivalen a dinero; en el caso de las universidades, al momento de solicitar el servicio, debe presentar fundamentos sólidos, para demostrar que sus proyectos son sobresalientes, y ser específicos en cada requisito de la solicitud.

LINKOGRAFÍA

A – C

Amazon Web Services, Inc. (2015). Obtenido de *amazon web services: Soluciones*. Recuperado el 9 de Diciembre del 2014 de http://aws.amazon.com/es/solutions/?nc2=h_ql

Ángel, Mieles. (20 de Enero de 2015). Obtenido de *Mialto: Introducción a la Arquitectura Cliente-Servidor(1)*. Recuperado el 30 de Enero del 2015 de <http://mialto.es/introduccion-a-la-arquitectura-cliente-servidor-1/>

Ángel, Mieles. (20 de Enero de 2015). Obtenido de *Mialto: : Introducción a la Arquitectura Cliente-Servidor(2)*. Recuperado el 30 de Enero del 2015 de <http://mialto.es/introduccion-a-la-arquitectura-cliente-servidor-2/>

Alcocer, A. (31 de Marzo de 2010). Obtenido de *SOCIETIC: Cloud Computing, características de las aplicaciones en cloud*. Recuperado el 5 de Diciembre del 2014. <http://www.societic.com/2010/03/cloud-computing-caracteristicas-de-las-aplicaciones-en-cloud/>

Asamblea Nacional República del Ecuador. (25 de Junio de 2013). Obtenido de la *Asamblea Nacional*. Recuperado el 19 de Enero del 2015 de http://www.asambleanacional.gob.ec/es/system/files/ley_organica_comunicacion.pdf

Asamblea Nacional República del Ecuador Obtenida de la *Ley Orgánica de Transparencia y Acceso a la Información Pública*. Recuperada el 29 de Enero del 2015 de http://www.cdc.gob.cl/wp-content/uploads/documentos/legislacion_internacional/ley_organica_de_acceso_a_la_informacion_en_ecuador.pdf

aprendaRedes.com. (2015). Obtenido de *AprendaRedes.com:Infraestructura*. Recuperado el 3 de Febrero del 2015 de <http://www.aprendaredes.com/blog/que-es-un-data-center-4/>

Brange, Á. (7 de Julio de 2013). Obtenido de *Alvaro Brange's: Machine Learning*. Recuperado el 10 de Diciembre del 2014 de

<http://brange.me/2013/07/07/apuntes-de-machine-learning/>

Cabeza, V. (6 de Marzo de 2012). Obtenido de *mcpro:IBM SmartCloud impulsa tu negocio a la nube*. Recuperado el 9 de Diciembre del 2014 de

<http://www.muycomputerpro.com/2012/03/06/ibm-smartcloud-negocio-nube>

Chiharu Murata, A. B.-R. (2015). Obtenido de <http://www.medigraphic.com/>. *Análisis discriminante para predecir el diagnóstico clínico de inmunodeficiencias primarias*.

Recuperado el 6 de Enero del 2015 de

<http://www.medigraphic.com/pdfs/revalেমex/ram-2015/ram152g.pdf>

CIBERNAT. (2014). Obtenido de *CIBERNAT:Computación en la Nube*. Recuperado el 22 de Diciembre del 2014 de <http://cibernat.com/articulos/computacion-en-la-nube>

CONCEPTO.DE. (2012-2015). Obtenido de *CONCEPTO.DE.Sistema Operativo*.

Recuperado el 4 de Diciembre del 2014 de <http://concepto.de/sistema-operativo/>

Culturación. (s.f.). Obtenido de *Culturación*. <http://culturacion.com/que-es-un-servidor-de-correo/>

D – G

definicion abc. (2007-2015). Obtenido de *definicion abc. Almacenamiento*.

Recuperado el 7 de Enero del 2015 de

<http://www.definicionabc.com/tecnologia/almacenamiento.php>

Definicion.de. (2008-2015). Obtenido de *Definicion.de. Inteligencia Artificial*.

Recuperado el 7 de Enero del 2015 de <http://definicion.de/inteligencia-artificial/>

definicionabc. (2015). Obtenido de *definicionabc. Aplicación*. Recuperado el 20 de Enero del 2015. <http://www.definicionabc.com/tecnologia/aplicacion.php>

dixxinet. (s.f.). Obtenido de *dixxinet. Proveedores de Servicio de Aplicaciones*. Recuperado el 18 de Febrero del 2015 de <http://www.dixxinet.com/application-service-provider-asp-proveedor-de-servicios-de-aplicaciones-psa>

EcuRed. (s.f.). Obtenido de *EcuRed*. Recuperado el 23 de Enero del 2015 de http://www.ecured.cu/index.php/Transmisi%C3%B3n_de_datos

EMC. (2015). Obtenido de *EMC: Nube Híbrida*. Recuperado el 4 de Febrero del 2015 de <http://mexico.emc.com/corporate/glossary/hybrid-cloud.htm>

Esthela R, (31 de Enero de 2008). Obtenido de *phpBB.Transmisión de Datos*. Recuperado el 12 de Enero del 2015 de <http://ofimatica.foroactivo.com/t8-definicion-de-transmision-de-datos>

Gartner. (2013). Obtenido de *Gartner.Glosario*. Recuperado el 19 de Enero del 2015 de <http://www.gartner.com/it-glossary/social-media>

GamerDic. (2013-2015). Obtenido de *GamerDic. Servidor de Escucha*. Recuperado el 12 de Enero del 2015 de <http://www.gamerdic.es/termino/servidor-de-escucha>

GCF Community Foundation International. (1998-2015). Obtenido de *GCFAprendeLibre. Aplicación Web*. Recuperado el 27 de Enero del 2015 de http://www.gcfaprendelibre.org/tecnologia/curso/informatica_basica/aplicaciones_web_y_todo_acerca_de_la_nube/1.do

Gonzalez, A. (30 de Julio de 2014). Obtenido de *Clevertask IT to Business Solutions. Conceptos Básicos de Machine Learning*. Recuperado el 18 de Febrero del 2015 de <http://www.clevertask.com/conceptos-basicos-machine-learning/>

Google. (2015). Obtenido de *Google Cloud Platform*. Recuperado el 28 de Enero del 2015 de <https://cloud.google.com/why-google/?hl=es>

Google. (2015). Obtenido de *Google Cloud Platform. Why Google Cloud Platform*. Recuperado el 12 de Febrero del 2015 de <https://cloud.google.com/appengine/docs/whatisgoogleappengine>

Google. (2015). Obtenido de *Google Cloud Platform.Documentation*. Recuperado el 9 de Febrero del 2015 de <https://cloud.google.com/compute/docs/>

Global Knowledge Training. (s.f.). Obtenido de *Global Knowledge.IBM Cloud & Smarter Infraestructure Training*. Recuperado el 29 de Enero de http://www.globalknowledge.net/mea-shared-content/documents/684697/886403/EMEA_IBM_CSI_Q1_2015_Web_KSA

Graterol, Y. (10 de Marzo de 2014). Obtenido de *Maestros del Web No SQL como el futuro de las bases de datos*. Recuperado el 14 de Enero del 2015 de <http://www.maestrosdelweb.com/nosql-como-el-futuro-de-las-bases-de-datos/>

GXportal. (2013). Obtenido de *Web Services*. Recuperado el 21 de Enero del 2015 de http://library.gxtechnical.com/gxdlsp/pub/genexus/internet/technicalpapers/web_services.htm

GZM, F. (2015). Obtenido de *Ducky Debug. Heroku*. Recuperado el 18 de Febrero del 2015 de <http://duckydebug.net/que-es-heroku/>

H – M

Herrera, S. M. (29 de Mayo de 2015). Obtenido de *Universidad Politécnica de Catalunya. Appnalyze*. Recuperado el 3 de Junio del 2015. <http://upcommons.upc.edu/pfc/bitstream/2099.1/26216/1/104615.pdf>

IDaccion. (2012). Obtenido de *IDacción. Cloud Computing*. Recuperado el 19 de Enero de <http://idnews.idaccion.com/ventajas-e-inconvenientes-del-cloud-computing-para-tu-empresa/>

INFORMÁTICA HOY. (2008). Obtenido de *INFORMÁTICAHOY. Servidores de impresión*. Recuperado el 11 de Diciembre del 2014 de <http://www.informatica-hoy.com.ar/aprender-informatica/Servidores-de-impresion.php>

InformaticaModerna.com. (2008-2015). Obtenido de *InformaticaModerna.com. Servidor*. Recuperado el 15 de enero del 2015 de <http://www.informaticamoderna.com/Servidor.htm>

Instituto Cervantes. (1997-2015). Obtenido de *Centro Virtual Cervantes* ©. *Metodología cuantitativa*. Recuperado el 10 de Marzo del 2015 de http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/metodologiacu ntitativa.htm

IBM (s.f). Obtenido de *IBM Knowledge Center*. Recuperado el 26 de Enero de http://www-01.ibm.com/support/knowledgecenter/SS8PJ7_7.0.0/org.eclipse.jst.ws.doc.user/concep ts/cwsstandards.html?lang=es

Interoute. (2013-2015). Obtenido de *Interoute from the ground to the cloud. PAAS*. Recuperado el 12 de Diciembre del 2014 de <http://www.interoute.es/what-paas>

Interoute. (2013-2015). Obtenido de *Interoute from the ground to the cloud. IAAS*. Recuperado el 12 de Diciembre del 2014 de <http://www.interoute.es/what-iaas>

json.org. (s.f.). Obtenido de *json.org*. Recuperado el 9 de Enero del 2015 de <http://json.org/json-es.html>

K35 IT Managers Group Team. (25 de Mayo de 2010). Obtenido de *K35. Nube privada*. Recuperado el 15 de Diciembre del 2014 de <http://grupok35.blogspot.com/2010/05/que-es-una-nube-privada-y-en-que-se.html>

Kezherashvili, B. (2011). Obtenido de *Administración de Sistemas operativos*. Recuperado el 15 de Diciembre del 2014 de http://www.adminso.es/recursos/Proyectos/PFM/2011_12/PFM_cloud_beka.pdf

Kezherashvili, B. (2011). Obtenido de *Computación en la Nube*. Recuperado el 15 de Diciembre del 2014 de http://www.adminso.es/recursos/Proyectos/PFM/2011_12/PFM_cloud_beka.pdf

KUMO. (s.f.). Obtenido de *KUMO. Cloud Computing*. Recuperado el 23 de Marzo del 2015 de <http://www.kumo.com.co/portal/index.php/pages/blog/175-ventajas-y-desventajas-del-cloud-computing>

Lara, K. R. (s.f.). Obtenido de *desarrolloweb.com. Aplicación de escritorio*. Recuperada el 16 de Diciembre del 2014 de <http://www.desarrolloweb.com/wiki/aplicacion-de-escritorio.html>

Leandro Alegsa, (2015). Obtenido de *ALEGSA.com.ar:Definición de Tecnología*.

Recuperado el 19 de Diciembre del 2014 de

<http://www.alegsa.com.ar/Dic/tecnologia.php>

Leandro Alegsa, (2015). Obtenido de *ALEGSA.com.ar: Diccionario de Informática y Tecnología. Recuperación de Datos*. Recuperado el 5 de Diciembre del 2014 de

<http://www.alegsa.com.ar/Dic/recuperacion%20de%20datos.php>

Leandro Alegsa, (2015). Obtenido de *ALEGSA.com.ar: Diccionario de Informática y Tecnología.Servidor de Aplicaciones*. Recuperado el 5 de Enero del 2015 de

<http://www.alegsa.com.ar/Dic/servidor%20de%20aplicaciones.php>

Leandro Alegsa. (2015). Obtenido de *ALEGSA.com.ar: Diccionario de Informática y Tecnología.XML*.Recuperado el 19 de Enero del 2015 de

<http://www.alegsa.com.ar/Dic/xml.php>

LibrosWeb.ec. (2015). Obtenido de *LIBROSWEB. Autorización*. Recuperado el 10 de Febrero del 2015 de http://librosweb.es/libro/symfony_2_x/capitulo_13/autorizacion.html

LibrosWeb.es. (2015). Obtenido de *LIBROSWEB.Seguridad*. Recuperado el 13 de Febrero del 2015 de http://librosweb.es/libro/symfony_2_4/capitulo_13.html

LibrosWeb.es. (2015). Obtenido de *LIBROSWEB*. Recuperado el 12 de Febrero del 2015 de

http://librosweb.es/libro/symfony_2_4/capitulo_13/como_funciona_la_seguridad_autenticacion_y_autorizacion.html

López, F. J. (2005). Grupos A Y B de Informática Bloque Específico de la Generalitat Valenciana Volumen 1. En F. J. López, *Grupos A Y B de Informática Bloque Específico de la Generalitat Valenciana Volumen 1* (págs. 43,44). España: MAD.

Martín, J. C. (5 de Octubre de 2007). Obtenido de *Blog del CIIN.SAAS*. Recuperado el 16 de Enero del 2015 de <http://geeks.ms/blogs/ciin/archive/2007/10/05/software-as-a-service-sas-191-qu-233-es.aspx>

MATERMAGAZINE. (s.f.). Obtenido de *Master Magazine. Cliente*. Recuperado el 2 de Diciembre del 2014 de <http://www.mastermagazine.info/termino/4292.php>

Masadelante.com. (2015). Obtenido de *masadelante.com. Base de Datos*. Recuperado el 17 de Enero del 2015 de <https://www.masadelante.com/faqs/base-de-datos>

Meltom Technologies. (2001-2015). Obtenido de *deGerencia.com: Tecnología de Información*. Recuperado el 3 de Diciembre del 2014 de http://www.degerencia.com/tema/tecnologia_de_informacion

Merchán, S. Y. (3 de Enero de 2013). Obtenido de *Programas Utilitarios II*. Recuperado el 22 de Diciembre del 2014 de http://programasutilitarios-unesum.blogspot.com/2013/01/v-behaviorurldefaultvml_3.html

Mesta, J. (3 de Octubre de 2011). Obtenido de *Geeks with blogs. Cloud Computing*. Recuperado el 8 de Diciembre del 2014 de <http://geekswithblogs.net/gotch/as/archive/2011/10/03/caracteristicas-esenciales-cloud-computing.aspx>

Microsoft. (2015). Obtenido de *Microsoft Azure*. Recuperado el 18 de Febrero del 2015 de <http://azure.microsoft.com/es-es/overview/what-is-azure/>

Miranda, A. T. (11 de Octubre de 2007). Obtenido de *Base de Datos. Bases de Datos Relacionales*. Recuperado el 18 de Diciembre del 2014 de <http://basededatosrelacionales.blogspot.com/2007/10/concepto-de-base-de-datos-relacionales.html>

misrespuestas.com. (2005-2015). Obtenido de *misrespuestas.com. Servidor Web*. Recuperado el 17 de Diciembre del 2015. <http://www.misrespuestas.com/que-es-un-servidor-web.html>

Molina, M. P. (13 de 04 de 2011). Obtenido de *e-coms.Bases de Datos*. Recuperado el 10 de Diciembre del 2014 de http://www.mariapinto.es/e-coms/bases_datos.htm

N – Q

Pallo, J. M. (7 de Noviembre de 2014). Obtenido de *El futuro everywhere. Machine Learning*. Recuperado el 20 de Febrero del 2015 de

<http://blogs.encamina.com/everyware/2014/11/07/un-poco-de-machine-learning-o-la-imitacion-binaria-de-el-agua-moja/>

Pasqualini, C. R. (2010). Obtenido de *Facultad de Ciencias de la Alimentación. Servidor de Archivo*. Recuperado el 16 de Diciembre del 2014 de <http://ayuda.fcal.uner.edu.ar/archivo/>

Prada, R. L. (2001). Obtenido de *Zaguán 2000. El método de investigación bibliográfica*. Recuperado el 20 de Mayo del 2015 de <http://www.oocities.org/zaguan2000/metodo.html>

ProspecNet S.L. . (2010). Obtenido de *ProspecNet. Cloud Computing*. Recuperado el 16 de Abril del 2015 de <http://www.prospecnet.com/CloudComputing.aspx>

PsicoActiva. (1998-2015). Obtenido de *PsicoActiva.com. Tecnología en la nube*. Recuperado el 9 de Diciembre del 2014 de <http://www.psicoactiva.com/arti/articulo.asp?SiteIdNo=138>

QODE. (31 de Octubre de 2012). Obtenido de *qodeblog. Aplicación Web*. Recuperado el 17 de Diciembre del 2014 de <http://qode.pro/blog/que-es-una-app/>

Quees.la. (2015). Obtenido de *Quees.la. Tecnología*. Recuperado el 8 de Diciembre del 2014 de <http://quees.la/tecnologia/>

R - Z - #

RECUPERADATA. (12 de Septiembre de 2012). Obtenido de *RECUPERADATA. Recuperación de información*. Recuperado 26 de Enero del 2015 de <http://www.recuperadata.com/noticias-sobre-recuperacion-de-datos.asp?IdContenido=376>

Reynoso, G. J. (12 de Octubre de 2010). Obtenido de *DDW. API's*. Recuperado el 22 de Enero del 2015 de <http://www.ddw.com.ar/blog/tecnologia-software-aplicaciones-y-servicios-web/331-que-es-y-para-que-sirve-una-api>

Rico, E. (5 de Febrero de 2015). Obtenido de *Otro Espacio Blog*. Recuperado el 7 de Febrero del 2015 de <https://otroespacioblog.wordpress.com/2015/02/05/servidores-para-videojuegos-que-debes-tener-en-cuenta/>

Rodriguez, M. (11 de Febrero de 2012). Obtenido de *Software Utilitario*. Recuperado el 9 de Enero del 2015 de <http://funcionesdeunsoftwareutilitario.blogspot.com/2012/02/un-utilitario-es-aquel-software-que.html>

Rodriguez, M. B. (s.f.). Obtenido de *Universidad del País Vasco. Web Services*. Recuperado el 16 de Enero del 2015 de <http://www.ehu.eus/mrodriguez/archivos/csharp/pdf/ServiciosWeb/WebServices.pdf>

Rouse, M. (Junio de 2014). Obtenido de *Search DataCenter. Autenticación Multifactor*. Recuperada el 2 de Febrero del 2015 de <http://searchdatacenter.techtarget.com/es/definicion/Autenticacion-multifactor-MFA>

Saffirio, M. (Febrero de 5 de 2006). Obtenido de *Tecnologías de Información y Gestión de Procesos de Negocios (BPM)*. Recuperado el 18 de Febrero del 2015 de <https://msaffirio.wordpress.com/2006/02/05/%C2%BFque-son-los-web-services/>

SCAS. (2010-2011). Obtenido de *SCAS. Manipulación de la Información*. Recuperado el 9 de Enero del 2015 de http://media-youth.org/index.php?option=com_content&view=article&id=15&Itemid=15&lang=es

Silvestrucci, G. M. (15 de Agosto de 2013). Obtenido de *Tecnologías de Información GTEK Honduras. Infraestructura*. Recuperado el 19 de Enero del 2015 de <http://gmsilvestrucci.blogspot.com/2013/08/definicion-de-la-infraestructura-de-ti.html?showComment=1431570868721#c4934519932252028878>

SVT Cloud. (6 de Octubre de 2012). Obtenido de *The Cloud Scurity Company. Seguridad como Servicio*. Recuperado el 23 de Febrero del 2015 de <http://blog.svtcloud.com/secaas-o-seguridad-como-servicio-no-solo-para-las-grandes-empresas/>

Turner, S. (14 de Abril de 2014). Obtenido de *Alojamiento Web. Bases de Datos*. Recuperado el 24 de Febrero del 2015 de <http://blog.iweb.com/es/2014/04/servidores-de-bases-de-datos/2487.html>

WeBlog, W. (s.f.). Obtenido de *WiWi WeBlog. Servidor de Aplicaciones*. Recuperado el 23 de Enero del 2015 de <https://wiwiloz.wordpress.com/servidor-de-aplicaciones/>

7G. (2013-2015). Obtenido de *Significados: Significados de Aprendizaje*. Recuperado el 4 de Diciembre del 2014 de <http://www.significados.com/aprendizaje/>

BIBLIOGRAFÍA

Max Welling Donald Bren, School of Information and Computer Science University of California. (April 21, 2010). A First Encounter with Machine Learning. California, Estados Unidos. Irvine.

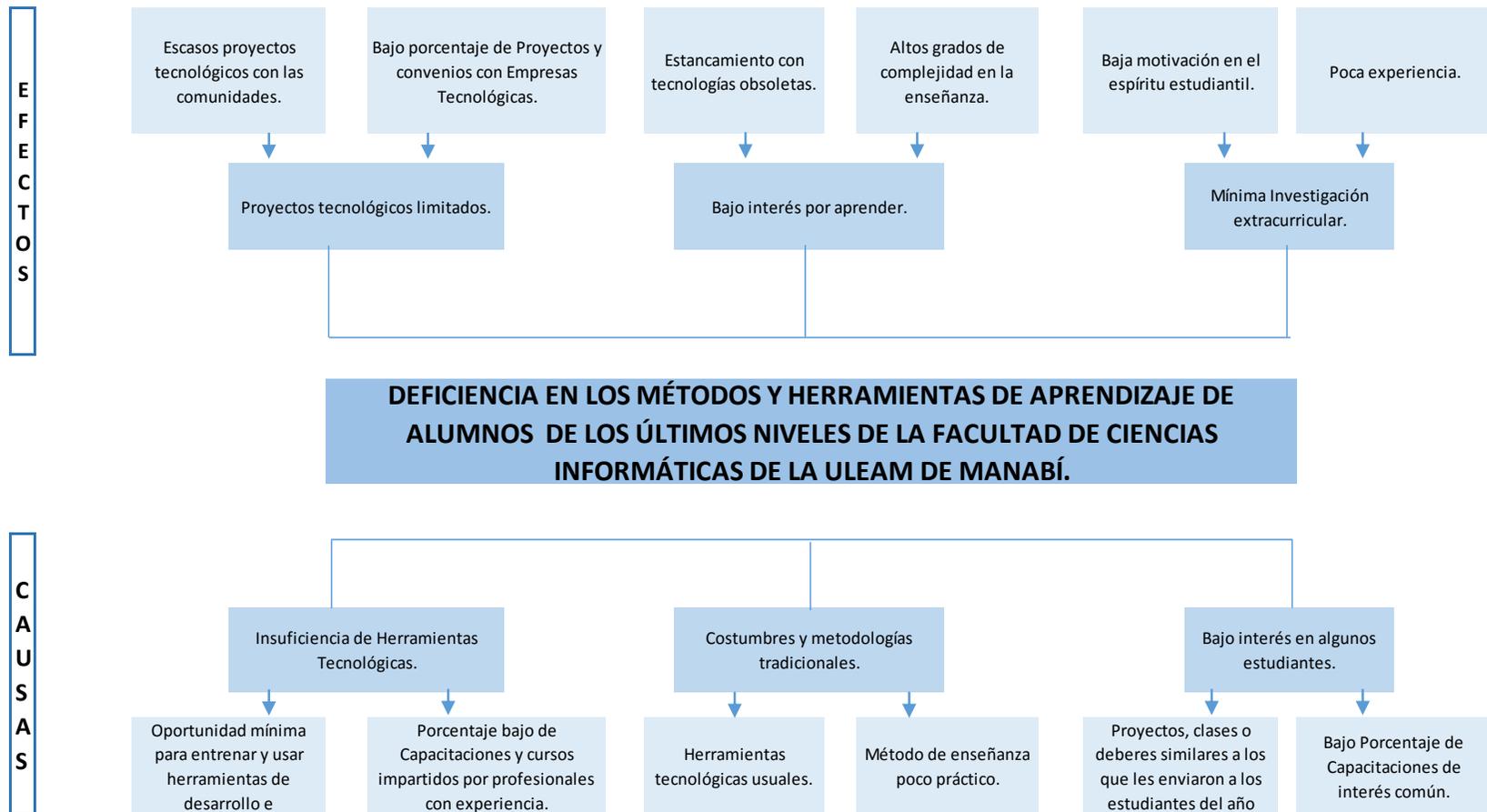
Alex Smola and S.V.N. Vishwanathan (2008) Introduction to Machine Learning. Madrid, Spain: Cambridge University Press 2008.

C. E. Rasmussen & C. K. I. Williams (2006) Gaussian Processes for Machine Learning, the MIT Press. London, England: Thomas Dietterich, Massachusetts Institute of Technology.

Stephen Marsland (2015) Machine Learning: An Algorithmic Perspective, Second Edition. U.S.A: Taylor & Francis Group.

ANEXOS

ANEXO 1. PLANTEAMIENTO DEL PROBLEMA



ANEXO 2. Encuesta No. 1

Encuesta realizada a los estudiantes de últimos niveles (7mo, 8vo, 9no) de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí, conocer sus competencias respecto a Machine Learning y Cloud Computing.

Pregunta No. 1: Para usted Cloud Computing es:

- a) Un paradigma que permite ofrecer servicios de computación a través de Internet.
- b) El uso de equipos de telecomunicaciones y computadoras para la transmisión, el procesamiento y el almacenamiento de datos.
- c) Una ciencia que estudia la creación y diseño de entidades capaces de resolver problemas cotidianos por sí mismas
- d) No sé

Pregunta No. 2: Para usted PAAS es:

- a) Plataforma como Servicio
- b) Puertos como Servicio
- c) Portal como Servicio
- d) No sé

Pregunta No.3. Para usted SAAS es:

- a) Shareware como Servicio
- b) Socket como Servicio
- c) Software como Servicio
- d) No sé

Pregunta No.4. Para usted IAAS es:

- a) Infraestructura como Servicio

- b) Interfaz como Servicio
- c) Inteligencia como Servicio
- d) No sé

Pregunta No. 5. ¿Conoce los tipos más conocidos de nube existentes en la actualidad? Encierre la opción escogida.

- a) Nube rápida
- b) Nube privada
- c) Nube comunitaria
- d) Nube distribuida
- e) Nube pública
- f) Nube híbrida
- g) Nube solitaria
- h) No sé

Pregunta No.6. ¿Ha escuchado sobre alguna de las siguientes plataformas de servicios? Encierre la opción escogida.

- a) MWA
- b) AWS
- c) Google Cloud Platform
- d) IBM SmartCloud
- e) Heroku
- f) No sé

Pregunta No. 7. Para usted Machine Learning es:

- a) Rama de la inteligencia artificial que busca desarrollar técnicas para que las computadoras aprendan.

- b) Razonamiento mediante lógica formal análogo al pensamiento abstracto inhumano.
- c) No sé

Pregunta No. 8. Sabe usted, ¿Cuál de las siguientes alternativas son tipos de Machine Learning? Encierre la opción escogida.

- a) Clustering
- b) Supervisado
- c) Parallel
- d) No supervisado
- e) No sé

Pregunta No. 9. ¿Bajo qué plataformas desarrolla aplicaciones móviles en la actualidad?

- a. Android
- b. IOS
- c. Windows Phone
- d. No desarrollo aplicaciones móviles en la actualidad

Pregunta No. 10. ¿Qué prefiere usted?

- a) Crear una aplicación web con herramientas como Wordpress
- b) Crear una aplicación web desarrollando en lenguajes como php, c#, entre otros
- c) Ninguna de las anteriores

ANEXO 3. Encuesta No. 2

Encuesta realizada a los estudiantes de últimos niveles (7mo, 8vo, 9no) de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí con el fin de conocer sus intereses.

Pregunta No.1. ¿Qué tan frecuente es la investigación que realiza en su casa respecto a lo impartido en clases?

- a) Todos los días
- b) días por semana
- c) 1 día por semana
- d) Ningún día

Pregunta No. 2. Al terminar su carrera como Ingeniero de Sistemas, le gustaría:

- a) Hacer una maestría, postgrado o especialización
- b) Hacer otra carrera
- c) Hacer cursos cortos sobre informática
- d) Ninguna de las anteriores

Pregunta No.3. Si hubiera una tecnología que le diera valor a sus aplicaciones, la aprendería.

- a) Si me incursionaría en el reto.
- b) Prefiero seguir usando las tecnologías que ya conozco.
- c) No desarrollo aplicaciones.
- d) No sé.

Pregunta No. 4. ¿Cuáles son las áreas en las que más destaca? Encierre la opción escogida.

- a. Desarrollo de Software
- b. Ingeniería de Software
- c. Admin. de Base de Datos
- d. Inteligencia Artificial
- e. Infraestructura en Redes
- f. Electrónica y Digitales
- g. No sé

Pregunta No. 5. Marque con X los dispositivos que posee:

- a. Celular básico o de gama baja
- b. Smartphone
- c. Tablet
- d. Laptop
- e. Desktop
- f. Wearables
- g. Ninguno de los anteriores

Pregunta No. 6. ¿Bajo qué denominación calificaría el acceso a Internet en general tomando en cuenta el de su hogar si lo tuviera, el de la facultad, o ambos?

- a) Alta disponibilidad
- b) Nivel medio
- c) Nivel bajo o casi nulo

ANEXO 4. Encuesta No. 3

Encuesta realizada a los docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí con el fin de conocer sus competencias respecto a Inteligencia Artificial, Machine Learning y Cloud Computing.

Pregunta No. 1 Para usted el aprendizaje es:

- a) Adquisición del conocimiento de algo por medio del estudio, el ejercicio o la experiencia, en especial de los conocimientos necesarios para aprender algún arte u oficio.
- b) Facultad del ser humano para comprender por medio de la razón la naturaleza, cualidades y relaciones de las cosas.
- c) No sé.

Pregunta No. 2 Si Machine Learning o Aprendizaje de Máquina es una rama de la Inteligencia Artificial, que desarrolla técnicas que permiten aprendizaje de las máquinas para hacer predicciones, diagnósticos, detecciones, etc. En la escala del 1 como bajo y 5 como máximo, determine la importancia, que cree usted tiene el aprendizaje del Machine Learning en la carrera de Ingeniería de Sistemas._____

Pregunta No. 3 Para usted Cloud Computing es:

- a) Un paradigma que permite ofrecer servicios de computación a través de Internet.
- b) El uso de equipos de telecomunicaciones y computadoras para la transmisión, el procesamiento y el almacenamiento de datos.
- c) No sé.

Pregunta No. 4 ¿Ha escuchado o ha usado alguna de las siguientes plataformas de servicios de Cloud Computing? Encierre en un círculo.

- a) Microsoft Azure.
- b) Google Cloud Platforms.
- c) IBM SmartCloud.
- d) AWS.
- e) No he escuchado, ni usado.

Pregunta No. 5 Señale las ramas de Inteligencia Artificial que usted ha escuchado o conoce.

- a) Redes Neuronales.
- b) Lógica Difusa.
- c) Algoritmos Genéticos.
- d) Robótica.
- e) Machine Learning.
- f) Otras.
- g) No he escuchado, ni conozco.

Pregunta No. 6 Según su criterio encierre una de las opciones, según lo que piensa es lo más importante para tener una predicción precisa en Machine Learning.

- a) Datos lo más normalizados posibles.
- b) Algoritmo eficiente.
- c) Gran cantidad de datos pasados.
- d) No sé

Pregunta No. 7 Sabe usted, ¿Cuál de las siguientes opciones son tipos del Machine Learning?

- a) Aprendizaje Supervisado.
- b) Parallel.
- c) Aprendizaje por posición.
- d) Transducción.
- e) No sé

Pregunta No. 8 Si ha hecho experimentos o aplicado el Machine Learning. ¿Bajo qué entorno para análisis de conocimiento lo ha hecho? Marque con x.

- a) Machine Learning en Microsoft Azure.
- b) Machine Learning en Weka.
- c) Machine Learning en Python.
- d) En otro entorno.
- e) No he aplicado el machine Learning aún.

ANEXO 5. Encuesta No. 4

Encuesta realizada a los estudiantes de últimos niveles (7mo, 8vo, 9no) y docentes de las áreas técnicas de la Facultad de Ciencias Informáticas de la Universidad Laica Eloy Alfaro de Manabí con el fin de conocer sus competencias respecto a Inteligencia Artificial, Machine Learning y Cloud Computing.

Pregunta No.1 Para usted Machine Learning es:

- a) Rama de la inteligencia artificial que busca desarrollar técnicas para que las computadoras aprendan.

- b) Razonamiento mediante lógica formal análogo al pensamiento abstracto inhumano.
- c) No sé

Pregunta No.2 ¿Cuál de las siguientes opciones cree usted son Tipos de Machine Learning?

- a) Clustering
- b) Supervisado
- c) Parallel
- d) No supervisado
- e) No sé

Pregunta No.3 Para usted Cloud Computing es:

- a) Un paradigma que permite ofrecer servicios de computación a través de Internet.
- b) El uso de equipos de telecomunicaciones y computadoras para la transmisión, el procesamiento y el almacenamiento de datos.
- c) Una ciencia que estudia la creación y diseño de entidades capaces de resolver problemas cotidianos por sí mismas No sé

Pregunta No.4 Señale la/las plataformas que conoce o ha escuchado:

- a) Microsoft Windows Azure
- b) AWS
- c) Google Cloud Platform
- d) IBM SmartCloud
- e) Heroku
- f) No sé

GLOSARIO

Daemon: es un tipo especial de proceso informático no interactivo, es decir, que se ejecuta en segundo plano en vez de ser controlado directamente por el usuario.

CORBA: es un estándar definido que permite que diversos componentes de software escritos en múltiples lenguajes de programación y que corren en diferentes computadoras, puedan trabajar juntos.

DCOM: (Modelo de Objetos de Componentes Distribuidos), tecnología Microsoft que desarrolla componentes de software para computadoras en red.

Sistema Distribuido: Colección de computadoras con distancias físicas, pero con conexión entre sí por medio de una red de comunicaciones.

Data Science: Extracción de conocimiento de grandes volúmenes de datos.