



UNIVERSIDAD TÉCNICA DE BABAHOYO

FACULTAD DE ADMINISTRACIÓN, FINANZAS E INFORMÁTICA

PROCESO DE TITULACIÓN

NOVIEMBRE 2021-ABRIL 2022

EXAMEN COMPLEXIVO DE GRADO O DE FIN DE CARRERA

PRUEBA PRÁCTICA

PREVIO A LA OBTENCIÓN DE TÍTULO DE INGENIERA EN SISTEMAS

TEMA:

**ANÁLISIS PARA LA APLICACIÓN DE LOS CLASIFICADORES EN EL
APRENDIZAJE AUTOMÁTICO**

EGRESADA:

NARCISA JANETH PEÑAFIEL BENÍTEZ

TUTOR:

ING. JOSÉ MEJÍA VITERI

AÑO 2022

INDICE

| | |
|--|----|
| RESUMEN | 3 |
| ABSTRACT | 4 |
| INTRODUCCION | 5 |
| DESARROLLO | 6 |
| El Machine Learning:..... | 8 |
| Figura 1 Regresión Logística para Clasificación | 12 |
| Figura 2: Arboles de decisión | 14 |
| Figura 3 Vectores de Soporte | 15 |
| Clasificadores de aprendizaje supervisados..... | 16 |
| Tabla 1. Comparativa de clasificadores de aprendizaje supervisados | 18 |
| CONCLUSIONES | 20 |
| REFERENCIAS | 21 |
| ANEXOS | 22 |

RESUMEN

Este documento es un caso de estudio que se orienta con la sublínea de investigación de la carrera de ingeniero de sistemas, y la sublínea es procesos de datos y telecomunicaciones”, tiene como objetivo el análisis de la aplicación de los clasificadores de aprendizaje automático y la forma en la que aportan a la ciencia y tecnología.

Es un documento relacionado netamente con el estudio de la inteligencia artificial implica y sus algoritmos de clasificación y recomienda unas formas de aplicación que pueden desplegarse y la naturaleza y el conjunto de datos disponibles para su aplicación.

Aquí se analizan los distintos métodos de aprendizaje automático existentes que tienen relación con la clasificación, desde las técnicas que sirven como base hasta los modelos más recientes, con el propósito de encontrar el mejor modelo que posibilite conocer la polaridad de una indicación.

Así mismo, es importante indicar que, se ha tomado importancia a la participación de profesionales de ingeniería en sistemas, conocedores de estos aspectos relacionados con Machine Learning de Inteligencia artificial, a los cuales se les ha hecho llegar una entrevista con instrumentos como formulaciones 4 preguntas extendidas para que con su opinión este caso de estudio se fortalezca además con el análisis de la autora.

Palabras Clave

Machine Learning, Python, Aprendizaje Automatizado, Inteligencia Artificial, Clasificadores de Aprendizaje.

ABSTRACT

This document is a case study that is oriented with the research subline of the systems engineer career, and the subline is data and telecommunications processes”, its objective is the analysis of the application of machine learning classifiers and the way in which they contribute to science and technology.

It is a document clearly related to the study of artificial intelligence and its classification algorithms, and recommends some forms of application that can be deployed and the nature and set of data available for its application.

Here, the different existing machine learning methods that are related to classification are analyzed, from the techniques that serve as a basis to the most recent models, with the purpose of finding the best model that makes it possible to know the polarity of an indication.

Likewise, it is important to indicate that importance has been given to the participation of systems engineering professionals, who are aware of these aspects related to Artificial Intelligence Machine Learning, to whom an interview has been sent with instruments such as formulations 4 questions extended so that with your opinion this case study is further strengthened with the author's analysis.

Key Words

Machine Learning, Python, Machine Learning, Artificial Intelligence, Learning Classifiers

INTRODUCCION

El aprendizaje automático se produce por medio de algoritmos. Un algoritmo no es más que una serie de pasos ordenados que se dan para realizar una tarea.

El objetivo del machine learning es crear un modelo que nos permita resolver una tarea dada. Luego se entrena el modelo usando gran cantidad de datos. El modelo aprende de estos datos y es capaz de hacer predicciones. Según la tarea que se quiera realizar, será más adecuado trabajar con un algoritmo u otro.

Aquí se estudian diferentes tipos de implementación de machine Learning, estas pueden clasificarse en tres categorías diferentes:

- Aprendizaje supervisado
- Aprendizaje no supervisado
- Aprendizaje de refuerzo según la naturaleza de los datos que recibe.

Con el aprendizaje supervisado, estos algoritmos hacen una labor de etiquetados, intentado y procesando datos para encontrar una función que, dadas las variables de entrada sea la adecuada para tener una salida adecuada. Este algoritmo es entrenado como un History Data y así procesa su aprendizaje a asignar la etiqueta para su salida adecuada con un nuevo valor.

En este estudio de caso se analizan los distintos métodos de aprendizaje automático existentes que tienen relación con la clasificación, desde las técnicas que sirven como base hasta los modelos más recientes, con el propósito de encontrar el mejor modelo que posibilite conocer la polaridad de una indicación.

Es importante estudiar los modelos de aprendizaje, pues hoy en día, con la completa penetración de Internet en la sociedad, esto es muy aprovechable para la ciencia y el desarrollo, así como útil para los negocios poder conocer lo que piensan sus usuarios de manera automática.

DESARROLLO

Este caso de estudio se orienta con la sublínea de investigación de la carrera de ingeniero de sistemas, pues la línea de investigación es “Comunicación y emprendimientos empresariales y tecnológicos, desarrollo de Sistemas de la información y la sublínea es procesos de datos y telecomunicaciones”.

El objetivo de este caso de estudio es analizar la aplicación de los clasificadores de aprendizaje automático y la forma en la que aportan a la ciencia y tecnología.

Estudiar la inteligencia artificial implica entender muchos algoritmos de clasificación que están disponibles en estos momentos para su utilización, sin embargo, no es posible concluir cuál es superior a otro o el más adecuado, pues depende de la aplicación que quiera dársele y la naturaleza y el conjunto de datos disponibles para su aplicación.

El aprendizaje sugiere reglas que se aprenden de forma secuencialmente utilizando los datos de entrenamiento uno a la vez, cada momento que se aprende una regla, se eliminan registros cubiertos por estas reglas, este proceso continúa con entrenamiento hasta que se cumplan condiciones determinadas, haciendo que un problema grande consiga un correcto algoritmo que permita aprendizaje automatizado.

Así mismo, es importante indicar que, el cerebro es el órgano más increíble del cuerpo humano, este establece la forma en que percibimos las imágenes, el sonido, el tacto, los olores y sabores, permite almacenar recuerdos, poder experimentar emociones e inclusive soñar, sin él, los seres humanos serían solamente organismos primitivos, incapaces de otra cosa que el más simple de los reflejos.

Es importante en este caso de estudio incorporar elementos teóricos que permitan enriquecer y fundamentar con bibliografías validas que brinden un mejor soporte, por lo que se tiene:

No es nada fácil definir el concepto de Inteligencia artificial. La informática la estudia la IA centrándose en las máquinas que exhiben una inteligencia humana o en el desarrollo de software. (Gamboa, 2018)

Los principales objetivos de la IA incluyen la capacidad de manipular y mover objetos, la presentación de conocimiento, la planificación, el procesamiento del lenguaje natural (NLP), la percepción, el aprendizaje, la inducción y el razonamiento. A largo plazo los objetivos incluyen la inteligencia general (a nivel humano), la inteligencia social y el logro de la creatividad. (Julio Ponce Gallegos, 2019)

Según (Muñoz, López, & Caicedo, 2008), Se puede abordar el problema de Inteligencia artificial y distinguir cuatro enfoques distintos.

Sistemas que se comportan como humanos. Desarrollar máquinas capaces que realicen funciones sería la idea aquí para las cuales un humano inteligente se requeriría. La famosa prueba de Turing la podemos encontrar Dentro de este enfoque. dónde la máquina debería poseer las siguientes capacidades para poder superar esta prueba. (Fredy Adolfo Forero Villa, 2020)

El comportamiento racional de algunos sistemas: Dentro de este enfoque es el diseño de agentes inteligentes o racional estos son aquellos que tienen como objetivo alcanzar el mejor resultado posible o el mejor posible en una situación de incertidumbre. (Rouhiainen, 2021)

Un punto importante es que la racionalidad perfecta o hacer siempre lo correcto no siempre es posible tarde o temprano no damos cuenta que esta situación más se presenta en los entornos complejos. Cuando la demanda computacional es elevada por lo cual la racionalidad limitada es hasta donde podremos llegar. (JUAN JESÚS ROMERO, 2017)

El realizar las inferencias de una manera correcta es lo que se busca con este enfoque, las habilidades necesarias para la prueba de Turing también son esenciales en este enfoque, básicamente porque tendremos que contar con la capacidad adecuada para exponer el conocimiento y la racionalización en base a él, ya que mediante estos se podrá alcanzar una correcta decisión en las situaciones de varias gamas. Generar sentencias de una forma comprensible en lenguaje natural es necesario, ya que en oraciones en los enunciados que se presentan permiten el debido desenvolvimiento de los agentes en una sociedad compleja. (Julio Cesar Ponce, 2019)

El aprendizaje no se da exclusivamente por erudición sino por el saber el cómo funcionan el mundo. Permite idear las debidas estrategias de cómo moverse o manejarse en el mismo. (Stuar Rusel, 2019)

Dentro de la Inteligencia Artificial se pueden encontrar distintas ramas, entre las que se destacan:

El Machine Learning: Es el análisis con diferentes herramientas informáticas las mismas que se usan para mejorar la toma de decisiones futuras, este es el estudio o análisis de programas que permiten aprender de los datos procesados. Esta tecnología de aprendizaje tiene como objetivo inducir a partir de ejemplos, una regla desconocida donde esta pueda ser bien aplicada. Uno de los usos o ejemplos en la actualidad sobre Machine Learning que ha sido muy visible es su aplicación en el filtrado de spam. (ManagementSolution, 2019)

La tecnología ML, cuenta con una gama ampliada de aplicaciones, estas son de diagnóstico médicos, detección oportuna de fraudes en el uso de tarjetas de crédito, de aplicación al análisis de bolsas de valores, secuencias de ADN y sus clasificaciones, reconocimiento voz en diferentes dispositivos, robótica y además los motores de búsqueda que comúnmente se utilizan. (ManagementSolution, 2019)

Una gran parte de conjunto de algoritmos de Machine Learning proviene del Deep que para encontrar patrones de datos utiliza estructuras profundas de redes neuronales. Estos tipos de algoritmo son de sumo interés por su tasa de éxito en la solución de determinados tipos de problemas. (Padial, 2020)

El razonamiento probabilístico se encarga de todo proceso de aprendizaje lidiando con la incertidumbre inherente que se presentan. El problema que se presenta en la creación de Inteligencia Artificial podemos decir que es la forma de trabajar con información ruidosa o incompleta e incluso contradictoria. Estos algoritmos están conectados con la estadística bayesiana; una de las herramientas principales en la que se basan es en el teorema de Bayes. (Padial, 2020)

La idea de estos algoritmos genéticos es imitar a la Evolución; funcionan seleccionando individuos de una población de soluciones candidatas, y luego intentando producir nuevas generaciones de soluciones mejores que las anteriores una y otra vez hasta aproximarse a una solución perfecta. (Goldberg, 1989)

Unas de las ramas de la inteligencia artificial es el aprendizaje automático el cual se aplica en los desarrollos de proyectos ya sea por la necesidad que tenga una máquina, que realice la clasificación y no se lleve a cabo un jurado de psicólogos, para que un

ordenador realice una predicción de la orientación de un comunicado corto se lo debe entrenar. (Fogel, Owens, & Walsh, 1966)

La inteligencia artificial tiene una de las áreas que es el aprendizaje automático, es capaz de crear sistemas que aprenden de forma automatizadas, muy capaces en la mejora de entrega de resultados. La ciencia informática ha supuesto un desarrollo hacia la nube con una relación de costo beneficio muy beneficiosa para aplicar técnicas de inteligencia artificial en la empresa. (Weiner, 1988)

Los algoritmos de IA, además son estadísticos, quiere decir que estos se basan en el conocimiento y la experiencia obtenida a partir de un indicio de conocimiento importante de información valida, esto le sirve como entrenamiento y aprendizaje. (Sandoval., 2019)

Es importante exponer, que este caso de estudio además ha sido fortalecido con un análisis comparativo de las entrevistas a expertos conocidos de la ingeniería en sistemas y que ejercen la profesión en diferentes ámbitos, los cuales han respondido basándose en sus experiencias.

En relación a: En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automáticos y cuáles han sido las razones. Se ha indicado que, en redes sociales y la medicina; describiendo y analizando de forma concluyente lo que los profesionales entrevistados han manifestado se tiene:

El proveer de herramientas informáticas con procedimientos que les permita procesar todos los datos de información médica sanitaria, sobre todo los que puedan provenir de Historias Clínicas, así mismo, emitir diagnósticos certeros que logren ofrecer datos predictivos sobre algunas enfermedades, de esta manera llegar a conocer el porcentaje de efectividad que tendría un tratamiento médico en un perfil concreto relacionado con algún paciente.

La utilización de algoritmos de Aprendizaje Automático, en los sistemas de diagnóstico, estos con la ayuda de la analítica, puede conllevar a mejoras importantes en la predicción de muchas enfermedades, para poder funcionar adecuadamente, el Machine Learning por lo general necesita enormes cantidades de datos, en tal sentido su funcionamiento y despliegue está íntimamente vinculado con otra nueva tecnología como

el Big Data; afortunadamente, muchos sistemas de salud, han optado por la digitalización. por tanto, el historial clínico en general de los pacientes y sus tratamientos ya están en formato digital, de esta manera el sistema puede alimentarse de ellos para encontrar datos útiles y ejercer la analítica y medicina predictiva.

En relación a: Indique algún escenario apropiado para la utilización de Machine Learning. Se ha indicado en Sugerencia de películas: Si alguna vez ha visto una película Alien y ET, y se desea saber si le puede interesar a una persona ver la película Parque Jurasico, con suficiente información y datos, a un clasificador le podría inferir de que si es probable que a una persona usuaria de Netflix por ejemplo esté en el mismo conjunto de preferencia de películas de los que se han interesado ya anteriormente por Jurassic Park.

También, se ha indicado referencias relacionadas con Sugerencias de compras: Un sitio web bien puede sugerir determinados productos evaluando los que anteriormente alguien ya ha comprado, a partir de la comparativa de sus compras con las de otros compradores, el clasificador le podría sugerir la compra de productos similares a los que usted ha adquirido.

Reconocimiento facial, esto tiene relación con el aprendizaje de tener algunos elementos del rostro previamente almacenados y haciendo que de alguna forma se registre o se entrene y se almacene en una base de datos, los rasgos identificativos de un rostro, para no solo detectar, sino también identificar a una persona.

En relación a: Cuales serían las secuencias o los pasos para poder utilizar el aprendizaje automático

1. Recolección de Datos: En cada situación que se desea una solución, se debe averiguar y extraer datos que se utilizarán para alimentar a una máquina.
2. Preparación de los datos: Es importante realizar visualizaciones de nuestros datos, para revisar y editar si es que hubieren correlaciones entre las distintas características, separar y organizar.
3. Elegir un modelo adecuado: Existen modelos diferentes que pueden elegirse de acuerdo al objetivo que se tenga, esto es: clasificación, regresión lineal, predicción, clustering, bayesiano, etc.
4. Entrenar nuestra máquina: Se utiliza un set de datos de entrenamiento para ejecutarla y se debe ver una mejora incremental (para las predicciones). Recordar inicializar los “pesos” del modelo de forma aleatoria.

5. Evaluación: Se debe comprobar que la máquina desarrollada un set de datos de Evaluación que cuenta con entradas que el modelo desconoce y en este caso hacer la verificación de la precisión del modelo ya entrenado; si es que la exactitud es menor o igual al 50%, es un modelo no útil; Si se alcanza un 90% o más se puede indicar que existe confianza positiva en los resultados otorgados por el modelo.
6. Parameter Tuning: Este tuneo es más un arte que una ciencia, y se irá mejorando a medida que se vaya experimentando, ya que suele haber muchos parámetros que se van ajustando y al combinarlos se pueden disparar todas las opciones buscadas.
7. Predicción o Inferencia: Ya listos para utilizar el modelo de Aprendizaje Automático, con nueva información y se da paso a la predicción e inferencia de resultados.

En relación a: Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia.

De las más conocidas son las librerías de Python, Matplotlib es una la librería gráfica de Python, se puede usar para generar gráficos de calidad necesaria para publicarlas tanto en papel como digitalmente, esto es: series temporales, histogramas, espectros de potencia, diagramas de barras, diagramas de errores, etc.

Scikit-learn es una librería de Python también, usada para Machine Learning y Análisis de Datos. Es basada en NumPy, SciPy y Matplotlib; de las ventajas principales se tiene, su facilidad de uso y la gran cantidad de técnicas de aprendizaje automático que permite implementar.

Análisis acerca de la clasificación supervisada, esta es una tarea que con más frecuencia es llevada a cabo por los denominados Sistemas de Inteligencia Artificial. Por lo que, se tiene un número enorme de paradigmas que se han desarrollado y han sido potenciados por la estadística (Regresión Logística, Análisis Discriminante) o por la Inteligencia Artificial (Redes Neuronales, Inducción de Reglas, Árboles de Decisión, Redes Bayesianas) capaces de realizar tareas propias de clasificación.

En cuanto a la **Regresión Logística para Clasificación**, según (José Martínez Heras), La regresión logística es una técnica de aprendizaje automático sumamente sencilla, los resultados son interpretables y es usada frecuentemente en la industria. Funciona perfectamente bien cuando hay muchísimos datos y las interrelaciones entre ellos no son muy complejas. (Gauchía & Gómez Adrian, 2019)

La Regresión Logística es una técnica de aprendizaje automático para clasificación, es como una red neuronal en miniatura, es así que, la regresión logística, se trata de una red neuronal con exactamente una neurona.

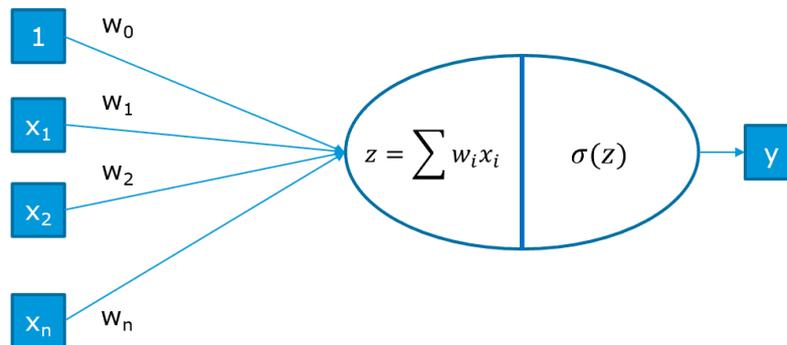


Figura 1 Regresión Logística para Clasificación

Fuente: José Martínez Heras

En la Figura 1, valores de x son correspondientes a los distintos atributos de un problema, por ejemplo, si se desea saber si un correo electrónico es deseado o no deseado (spam), los valores de x corresponderían a cuántas veces aparece cada palabra en un texto, la predicción y sería la probabilidad de que el correo sea spam (no deseado).

La regresión logística es un método de clasificación simple y efectivo para muchos problemas, es sencillo de interpretar y se puede inspeccionar qué factores contribuyen para obtener un tipo específico de resultados.

Cuando se realiza un análisis de noticias de portada, se utiliza regresión logística para predecir la probabilidad de que una noticia tuviese muchos me gusta o muchas visitas

En el campo del aprendizaje automático, el objetivo de la clasificación es recopilar objetos con características similares. Un clasificador lineal toma decisiones de clasificación a través de combinaciones lineales de características para lograr este objetivo. La característica de un objeto generalmente se describe como un valor de característica, y en un vector, se describe como un vector de característica, a esto se le puede conocer como Clasificador lineal y discriminante de Fisher.

Así mismo, una **Red Neuronal Artificial** (RNA) es un modelo matemático inspirado en el comportamiento biológico de una neurona, y en cómo estas se organizan formando una estructura cerebral. Estas intentan aprender mediante ensayos repetidos como organizarse mejor a sí mismas para lograr maximizar la predicción.

Las redes neuronales se hicieron operativas por primera en los 50. Rosenblat F. (1958) creó el perceptrón, este es un algoritmo avanzado de reconocimiento de patrones, basado en una red para aprendizaje usando una simple suma y la resta, luego el proceso evolucionó en los 60 y esta tecnología sufrió un estancamiento, luego en los 80 fue donde se logró el mayor desarrollo teórico. Son muchos los modelos de redes neuronales, el más utilizado es el algoritmo backpropagatio

Existen actualmente más de 40 paradigmas de redes neuronales artificiales. Se estima que tan sólo cuatro arquitecturas:

1. el modelo perceptrón multicapa (MLP),
2. los mapas autoorganizados de Kohonen, (SOFM),
3. el vector de cuantificación (LVQ) y
4. las redes de base radial (RBF)

Ejemplo de red neuronal: El perceptón, se conoce que el elemento básico de una red neuronal es un nodo, esta es la unidad de procesamiento que actúa en paralelo con otros nodos de la red. Es igual a las neuronas del cerebro de las personas, esto es acepta input y genera output, sus nodos aceptan input de otros nodos, la primera tarea del nodo es procesar datos de entrada creando valores de resumen que es la suma de todas las entradas multiplicadas aplicándole ponderaciones; este valor resumen se procesa a continuación mediante una función de activación para generar una salida que se envía al siguiente nodo del sistema.

Un ejemplo de modelo neuronal con nn entradas, consta de:

- Un conjunto de entradas x_1, \dots, x_n
- Los pesos sinápticos w_1, \dots, w_n , correspondientes a cada entrada.
- Una función de agregación, \sum .
- Una función de activación, ff.
- Una salida, Y.

Las entradas son el estímulo que la neurona artificial recibe del entorno que la rodea, y la salida es la respuesta a tal estímulo. La neurona puede adaptarse al medio circundante y aprender de él modificando el valor de sus pesos sinápticos, y por ello son

conocidos como los parámetros libres del modelo, ya que pueden ser modificados y adaptados para realizar una tarea determinada.

En este modelo, la salida neuronal Y está dada por:

$$Y = f\left(\sum_{i=1}^n w_i x_i\right)$$

Aplicaciones: En el reconocimiento de patrones, el discriminante lineal de Fisher es un método discriminante lineal cuya intención es reflejar puntos de datos específicos en un espacio d -dimensional en un espacio $c-1$ dimensional, de tal modo que en este se encuentren diferentes tipos muestrales, las proyecciones de las imágenes deben separarse lo más posible, y los similares deben ser lo más compactos posible.

Los **árboles de decisión o clasificación** no son modelos estadísticos basados en la estimación de los parámetros de la ecuación propuesta, por lo tanto, no se debe estimar un modelo estadístico formal en estos casos, más bien son algoritmos para clasificar utilizando particiones sucesivas, estos son adecuados cuando existe un gran número de datos.

Son además un modelo surgido en el ámbito (Machine Learning) y de la Inteligencia Artificial, que partiendo de una base de datos, permite crear diagramas con construcciones lógicas que permiten resolver problemas.

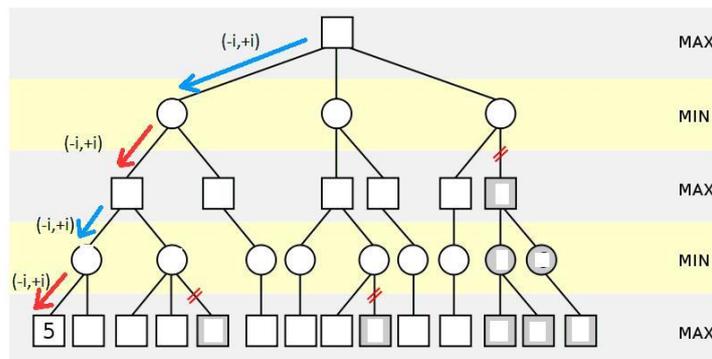


Figura 2: Árboles de decisión

Fuente: de Francisco Parra

En estos, sus reglas de parada tratan de preguntar si vale la pena seguir o no el proceso de crecimiento del árbol por una rama actual, a esto se le denominan reglas de preopda, ya que reducen su crecimiento mientras se está construyendo el árbol:

Pureza de nodo. Si el nodo solo contiene ejemplos o registros de una única clase se decide que la construcción del árbol ya ha finalizado.

Cota de profundidad. Previamente a la construcción se fija una cota que nos marque la profundidad del árbol, cuando se alcanza se detiene el proceso.

Umbral de soporte. Se especifica un número de ejemplos mínimo para los nodos, y cuando se encuentre un nodo con ejemplos por debajo del mínimo se para el proceso, ya que no consideramos fiable una clasificación abalada con menos de ese número mínimo de ejemplos.

Las máquinas de soporte vectorial (SVM), estas máquinas de vectores de soporte, poseen una fundamentación matemática pura dentro de la estadística del aprendizaje, a pesar de esto, su implementación básica tiene algunas limitaciones puesto que están diseñadas originalmente para problemas de clasificación binarios, y además tienen como limitante el que su algoritmo básico para entrenamientos les genera gran cantidad de vectores soporte, lo que provoca una lentitud en la clasificación, sin embargo son una poderosa herramienta que está destinada a labores de clasificación. Esta es muy usada en el campo de la neuroinformática.

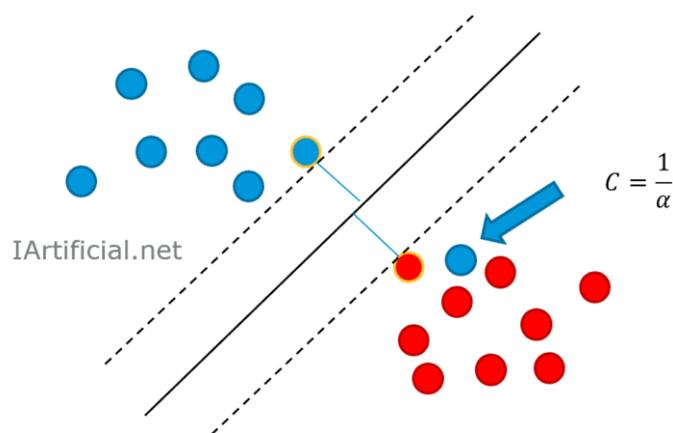


Figura 3 Vectores de Soporte

Fuente: Jose Martinez Heras

Este algoritmo se basa en el mapeo de puntos para su entrenamiento a un espacio vectorial, considerando una dimensión mayor, con esto construye hiperplanos en múltiples dimensiones espaciales que le permite separar los puntos en sus clases respectivas y así lograr clasificar puntos nuevos de acuerdo a una referencia de ubicación respecto del hiperplano de separación. (Delgado Castillo Duniel, 2016)

Clasificadores de aprendizaje supervisados

| Regresión Logística para Clasificación | Red Neuronal Artificial (RNA) | árboles de decisión o clasificación | Máquinas de soporte vectorial |
|--|--|--|--|
| VENTAJAS | | | |
| <p>Se pueden modelar relaciones complejas, no lineales. Robusto al ruido, esto se debe a que maximizan los márgenes. La regresión logística es más fácil de implementar, interpretar y muy eficiente de entrenar. No hace suposiciones sobre distribuciones de clases en el espacio de características. No hace suposiciones sobre distribuciones de clases en el espacio de características. No solo proporciona una medida de cuán apropiado es un predictor (tamaño del coeficiente), sino también su dirección de asociación (positiva o negativa). Es muy rápido para clasificar registros desconocidos. Buena precisión para muchos conjuntos de datos simples y funciona bien cuando el conjunto de datos es separable linealmente. Puede interpretar los coeficientes del modelo como indicadores de importancia de características.</p> | <p>Se basan en la estructura del sistema nervioso del cerebro. Aprendizaje, pues tienen la habilidad de aprender mediante una etapa que se llama etapa de aprendizaje. Esta consiste en proporcionar a la RNA datos como entrada a su vez que se le indica cuál es la salida (respuesta) esperada. Auto organización pues crea su propia representación de la información en su interior, descargando al usuario de esto. Tolerancia a fallos, Debido a que una RNA almacena la información de forma redundante, ésta puede seguir respondiendo de manera aceptable aun si se daña parcialmente. Flexibilidad, porque puede manejar cambios no importantes en la información de entrada, como señales con ruido u otros cambios en la entrada Tiempo real La estructura de una RNA es paralela, por lo</p> | <p>Muy fácil de interpretar y entender. Rápido. Robusto al ruido y valores perdidos. Preciso Excelente para aprender relaciones complejas, altamente no lineales. Por lo general, pueden lograr un rendimiento bastante alto, mejor que la regresión polinomial. Son simples de entender y de interpretar Si el árbol no es excesivamente grande, puede visualizarse No requiere una preparación de los datos demasiado exigente (aunque la implementación de Scikit-Learn no soporta valores nulos) Se puede trabajar tanto con variables cuantitativas como cualitativas Utiliza un modelo de caja blanca: la respuesta del algoritmo es fácilmente justificable a partir de la lógica booleana implementada en él</p> | <p>Eficaz en espacios de grandes dimensiones. Aun eficiente en casos donde su dimensión es mayor que al número de muestra. Usa un subconjunto de puntos para el entrenamiento en la función de decisión bien llamada vectores de soporte, esta es eficiente además en memoria. Versátil: Se especifican algunas funciones de núcleo para la función de decisión. Se proporcionan kernels similares, y también es posible especificarlos de forma personalizados.</p> |

| | | | |
|---|--|--|--|
| <p>Es menos propensa a sobreajustar, pero puede sobreajustarse en conjuntos de datos de alta dimensión.</p> | <p>cual, si esto es implementado con computadoras o en dispositivos electrónicos especiales, se pueden obtener respuestas en tiempo real. Es un modelo de vanguardia que capturan de una forma óptima y efectiva características</p> | | |
|---|--|--|--|

| DESVENTAJAS | | | |
|---|--|---|---|
| <p>Necesidad de seleccionar una buena función de kernel. Los parámetros del modelo son difíciles de interpretar. Requiere memoria significativa y poder de procesamiento. Cuando se tiene muchos datos toma demasiado tiempo para entrenar. Si el número de observaciones es menor que el número de características, no se debe utilizar la regresión logística; de lo contrario, puede provocar un sobreajuste. Construye límites lineales. La principal limitación de la regresión logística es el supuesto de linealidad entre la variable dependiente y las variables independientes. Solo se puede usar para predecir funciones discretas. Por tanto, la variable dependiente de</p> | <p>Complejidad de aprendizaje para grandes tareas. Tiempo de aprendizaje elevado. No permite interpretar lo que se ha aprendido, la red por si sola proporciona una salida, un número, que no puede ser interpretado por ella misma, sino que se requiere de la intervención del programador y de la aplicación en si Elevada cantidad de datos para el entrenamiento, La falta de reglas definitorias que ayuden a realizar una red para un problema dado. No son fácilmente explicables, conocer las reglas o motivos por los que la red devuelve esos resultados no suele ser fácil y precisa de otras analíticas. Tiempo de aprendizaje elevado que depende de dos</p> | <p>Los árboles complejos son difíciles de interpretar. Es posible la duplicación dentro del mismo subárbol. En ocasiones no es utilizado por ser un algoritmo tan sencillo y no tan poderoso para datos complejos. Los aprendices de árbol de decisión tienden al sobre entrenamiento, especialmente cuando el número de características predictivas es alto Son inestables: cualquier pequeño cambio en los datos de entrada puede suponer un árbol de decisión completamente diferente. No se puede garantizar que el árbol generado sea el óptimo. Hay conceptos que no son fácilmente aprendibles pues los árboles de decisión no son capaces de expresarlos con facilidad (como el operador XOR). Los aprendices crean árboles sesgados si hay</p> | <p>Solo están diseñadas originalmente para problemas de clasificación binarios es decir dos tipos Su algoritmo básico para entrenamientos les genera gran cantidad de vectores de soporte Lentitud en la clasificación. Si el número de características es mucho mayor que el número de muestras, evite el exceso de ajuste al elegir las funciones del Kernel y el término de regularización es crucial. Los SVMs no proporcionan directamente estimaciones de probabilidad, éstas se calculan utilizando una validación cruzada quíntuple (véase Partituras y probabilidades, más abajo).</p> |

| | | | |
|--|---|--|---|
| <p>Regresión logística está ligada al conjunto de números discretos. Los problemas no lineales no se pueden resolver con regresión logística porque tiene una superficie de decisión lineal. Los datos linealmente separables rara vez se encuentran en escenarios del mundo real. La regresión logística requiere una multicolinealidad promedio o nula entre variables independientes.</p> | <p>factores: A) Si se incrementa la cantidad de patrones a identificar o clasificar, y B) Si se requiere mayor flexibilidad o capacidad de adaptación de la red neuronal para reconocer patrones que sean sumamente parecidos, se deberá invertir más tiempo en lograr que la red converja a valores de pesos que representen lo que se quiera enseñar.</p> | <p>clases dominantes, por lo que se recomienda balancear el conjunto de datos antes de entrenar el aprendiz</p> | |
| UTILIDAD O APLICABILIDAD | | | |
| <p>Clasificación de texto e imágenes. Reconocimiento de escritura a mano. Para los casos en los que se desea predecir la presencia o ausencia de una característica o resultado según los valores de un conjunto de predictores</p> | <p>En el reconocimiento de patrones Predicción de sucesos y simulaciones Procesamiento de datos y modelización Para problemas de clasificación, como, por ejemplo, obtener la probabilidad de que un cliente te compre frente a que no te compre).</p> | <p>Diagnóstico médico. Análisis de riesgo crediticio. Un árbol de decisión Permite que un individuo o una organización comparen posibles acciones entre sí según sus costos, probabilidades y beneficios.</p> | <p>Reconocimiento óptico de caracteres Detección de caras para que las cámaras digitales enfoquen correctamente Filtros de spam para correo electrónico Reconocimiento de imágenes a bordo de satélites, para saber qué partes de una foto esta con nubes, tierra, hielo, agua etc.)</p> |

Tabla 1. Comparativa de clasificadores de aprendizaje supervisados

Este proceso investigativo está apegado a los estándares y normativas de la carrera de sistemas, con un enfoque bibliográfico cualitativo, consultado de bases teóricas sólidas y confiables donde se explora la producción de la comunidad académica sobre los clasificadores de aprendizaje de Machine Learnig de la Inteligencia Artificial, las herramientas y fuentes que tiene este trabajo investigativo son libros y artículos científicos elaborados por especialistas en IA, que permiten enriquecer este caso de estudio.

De este análisis bibliográfico, donde se consultó acerca de temas y conceptos de Inteligencia artificial, Machine Learning y sus Clasificadores de Aprendizaje, cuáles eran los más usados y más eficientes, así mismo de revistas científicas se pudo extraer conocimiento experimentado sobretodo acerca de los clasificadores de aprendizaje y su aplicabilidad en diferentes ámbitos, se presenta de forma sintetizada en la Tabla 1, donde se reflejan atributos importantes en pros y contras, así como sus usos, de cada tipo o técnica algorítmica de clasificadores de aprendizaje artificial.

Así mismo es importante indicar que esta investigación utiliza técnicas de entrevistas a profesionales del medio en la ciudad, como recogida de datos y por lo tanto como acción metodológica se podría indicar que este de carácter inductivo, porque va cada análisis puntual a una idea de resultado general, estas entrevistas permitieron que estas opiniones sean un aval importante que permitan validar este caso de estudio y tener recomendaciones técnicas de como por ejemplo iniciar a utilizar estos algoritmos y aplicarlos de forma más fácil con librerías como Scikit-learn, Keras, nltk y shap de python, usada para Machine Learning y Análisis de Datos.

La propuesta resultante de esta investigación es el procesamiento de datos de índole estadístico, de procesamiento de imágenes o de análisis de datos con bases de datos, apegado o con la utilización de estos algoritmos que están más a la mano y que cuentan con una gama de herramientas en varios lenguajes de programación como Python a través de librerías de gran disponibilidad para la comunidad científica y de la ingeniería en sistemas, con esto se podrá beneficiar de la exactitud y precisión de los resultados, así como su sensibilidad de las fuentes de datos aplicando las técnicas de aprendizaje automático que son de gran importancia para la analítica de datos en todo sentido y pueden ser muy útiles para infinidades de aplicaciones en futuros estudios y de fuente informativa para futuros profesionales.

El objetivo principal de este caso de estudio es encontrar los métodos adecuados para para la clasificación automática y el aprendizaje y cuáles de ellos tiene los mejores resultados, por lo que se han considerado:

Regresión Logística para Clasificación Red Neuronal Artificial (RNA)
árboles de decisión o clasificación Máquinas de soporte vectorial.

Además, se ha fortalecido con un estudio de los rasgos de estos mencionados métodos que permitan lograr una correcta selección y donde los clasificadores puedan mejorar su desempeño.

CONCLUSIONES

En la inteligencia artificial existen muchos algoritmos de clasificación y están disponibles en estos momentos para su utilización, pero no es posible concluir cuál es superior a otro o el más adecuado, todo depende de la aplicación y la naturaleza del conjunto de datos disponibles.

En el contexto del aprendizaje automático, y el problema de su clasificación, hay que considerar la supervisión donde depende netamente del algoritmo que aprende de un conjunto de datos, y este aprendizaje es utilizado para clasificar nuevas observaciones y resultados.

La exactitud y precisión de los datos, así como su sensibilidad por las técnicas de las técnicas de aprendizaje automático usadas son de gran importancia para el análisis de información en todo sentido y pueden ser muy útiles para infinidad de aplicaciones en futuros estudios.

Los métodos de clasificación encontrados o recogidos de varias fuentes en este caso de estudio, permiten evidenciar los mejores algoritmos y de más fácil acceso que cuentan además con una fuente clara de librerías en lenguajes fáciles de implementar como Python aportan de forma considerable a los conocimientos de los estudiantes de ingeniería de sistemas y a los desarrolladores y analistas de datos a proyectar mejores soluciones apegados al campo de la inteligencia artificial.

REFERENCIAS

- Fogel, L. J., Owens, A. J., & Walsh, M. J. (1966). *Artificial intelligence through simulated evolution*. John Wiley & Sons Inc.
- Fredy Adolfo Forero Villa, C. M. (2020). *Inteligencia artificial*. Bogotá, Colombia: <https://cerlalc.org/>.
- Gamboa, H. A. (2018). *INTELIGENCIA ARTIFICIAL*. Quito: www.researchgate.net.
- Gauchía, J. Z., & Gómez Adrian, J. A. (2019). *Análisis y comparación de algoritmos de detección*. España: etsin.
- Goldberg, D. E. (1989). *Genetic Algorithms in Search, Optimization and Machine Learning*. USA: Addison-Wesley Professional; 13th ed. edición.
- JUAN JESÚS ROMERO, C. D. (2017). *INTELIGENCIA ARTIFICIAL Y COMPUTACION AVANZADA*. QUITO: <https://cdv.dei.uc.pt/>.
- Julio Cesar Ponce, A. T. (2019). *Inteligencia Artificial*. PERU: <https://www.researchgate.net/>.
- Julio Ponce Gallegos, A. T. (2019). *Inteligencia Artificial*. Perú: es.scribd.
- ManagementSolution. (2019). *Machina Learning Una Pieza Clave en la Tranformacion*. España: ManagementSolution.
- Muñoz, M. A., López, J., & Caicedo, E. (2008). Inteligencia de enjambres: sociedades para la solución de. *Ingenieria e Investigacion*, 121.
- Padial, J. (2020). *TÉCNICAS DE PROGRAMACIÓN «DEEP LEARNING»: ¿Simulacro o realización artificial de la inteligencia?* Malaga: Universidad de Málaga.
- Rouhiainen, L. (2021). *Inteligencia artificial para los negocios*. <https://www.amazon.es/>.
- Sandoval., L. J. (2019). ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS. *ITCA FEPADE*, 36-40.
- Stuar Rusel, P. N. (2019). *Inteligencia Artificial Un Enfoque Morderno*. Mexico: PEARSON EDUCACIÓN, S.A.
- Weiner, N. (1988). *CIBERNETICA Y SOCIEDAD*. Sudamericana S.A.

ANEXOS

FORMULARIO DE ENTREVISTAS A EXPERTOS

Entrevista relacionada con:

Análisis para aplicación de los clasificadores de aprendizaje automático

Área: Inteligencia Artificial

PROFESIONAL (Nombres):

LUGAR DE TRABAJO:

- 1) En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automático y cuáles han sido las razones.**

- 2) Indique algún escenario apropiado para la utilización de Machine Learning**

- 3) Cuales serían las secuencias o los pasos para poder utilizar el aprendizaje automático**

- 4) Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia**

Respuestas de los Formularios:

FORMULARIO DE ENTREVISTAS A EXPERTOS

Entrevista relacionada con:

Análisis para aplicación de los clasificadores de aprendizaje automático

Área: Inteligencia Artificial

PROFESIONAL (Nombres): Freddy Moran

LUGAR DE TRABAJO:

1) En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automáticos y cuáles han sido las razones.

El aprendizaje automático se produce por medio de algoritmos. Un algoritmo no es más que una serie de pasos ordenados que se dan para realizar una tarea.

Vehículos inteligentes: ya veremos coches inteligentes en las carreteras. Gracias al aprendizaje automático, estos vehículos podrán ajustar la configuración interna (temperatura, música, inclinación del respaldo, etc.) de acuerdo a las preferencias del conductor e, incluso, mover el volante solos para reaccionar al entorno.

Redes sociales: Twitter, por ejemplo, se sirve de algoritmos de Machine Learning para reducir en gran medida el spam publicado en esta red social mientras que Facebook, a su vez, lo utiliza para detectar tanto noticias falsas como contenidos no permitidos en retransmisiones en directo que bloquea automáticamente.

Medicina: investigadores del Instituto de Tecnología de Massachusetts (MIT) ya utilizan el Machine Learning para detectar con mayor antelación el cáncer de mama, algo de vital importancia ya que su detección temprana aumenta las probabilidades de curación. Asimismo, también se utiliza con una alta eficacia para detectar neumonía y enfermedades de la retina que pueden provocar ceguera.

Ciberseguridad: los nuevos antivirus y motores de detección de malware ya se sirven del aprendizaje automático para potenciar el escaneado, acelerar la detección y mejorar la habilidad de reconocer anomalías.

2) Indique algún escenario apropiado para la utilización de Machine Learning

- Seguridad informática, diagnóstico de ataques, prevención de fraude online, detección de anomalías, etc.
- Reconocimiento de imágenes o patrones (facial, dactilar, objetos, voz, etc)
- Conducción autónoma, mediante algoritmos ~~deep learning~~: identificación de imágenes en tiempo real, detección de obstáculos y señales de tráfico, prevención de accidentes...
- Salud: evaluación automática de pruebas diagnósticas, robótica médica etc.
- Análisis de mercado de valores (predicciones financieras, evolución de mercados etc)

3) Cuales serían las secuencias o los pasos para poder utilizar el aprendizaje automático

Para lograr el éxito en la implementación del aprendizaje automatizado: identificar el problema o problemas que desea abordar, recopilar los datos, crear el modelo, implementarlo y supervisar su éxito.

4) Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia

FORMULARIO DE ENTREVISTAS A EXPERTOS

Entrevista relacionada con:

Análisis para aplicación de los clasificadores de aprendizaje automático

Área: Inteligencia Artificial

PROFESIONAL: Cristina Prieto

LUGAR DE TRABAJO: Babahoyo

1) En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automático y cuales han sido las razones.

Los clasificadores de aprendizaje se lo encuentran en algunas áreas como

Biológica (seres vivos y orgánico)

Biología computacional (algoritmo)

~~Biología~~ Bioinformática (desarrollo de herramientas informática)

Las razones es que cada una tiene su proceso en el que el ser humano desarrolla por medio de la inteligencia Artificial estudio necesarios para ayudar y aumentar las tecnologías para el beneficio de la humanidad

2) Indique algún escenario apropiado para la utilización de Machine Learning

En la salud

La seguridad ciudadana

La educación

3) Cuales serían las secuencias o los pasos para poder utilizar el aprendizaje automático

Para efectuar un buen aprendizaje automático debes identificar el problema en el que quieres trabajar, ya que pueden existir componentes que podrían afectar, recopilar los datos necesarios, formar un modelo de trabajo, implementarlo y supervisar su elaboración.

4) Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia

Existen algunos softwares, y plataformas que usan la inteligencia artificial como el Android en los teléfonos, las páginas web en este caso (Facebook, ~~instagram~~, lanzamientos de venta de artículos) que han catapultado como ayuda a la humanidad para mejorar su porvenir por medio del internet

FORMULARIO DE ENTREVISTAS A EXPERTOS

Entrevista relacionada con:

Análisis para aplicación de los clasificadores de aprendizaje automático

Área: Inteligencia Artificial

PROFESIONAL (Nombres): Adriana López Contreras

LUGAR DE TRABAJO: PC SOLUCIONES

1) En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automáticos y cuales han sido las razones.

En el área de salud

Un ejemplo de aprendizaje sería la construcción de un modelo de reingresos en hospitalización partiendo de un conjunto de datos previo de los que conocemos si el paciente reingresó o no

2) Indique algún escenario apropiado para la utilización de Machine Learning

Seguridad informática

3) Cuales serían las secuencias o los pasos para poder utilizar el aprendizaje automático

- Recursos de preparación de datos.
- Algoritmos – básicos y avanzados.
- Automatización y procesos iterativos.
- Escalabilidad.
- Modelado en conjunto

4) Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia

FORMULARIO DE ENTREVISTAS A EXPERTOS

Entrevista relacionada con:

Análisis para aplicación de los clasificadores de aprendizaje automático

Área: Inteligencia Artificial

PROFESIONAL (Nombres): HARRY SALTOS VITERI

LUGAR DE TRABAJO: UNIVERSIDAD TECNICA DE BABAHOYO

1) En qué área ha utilizado o ha visto la utilización de los clasificadores de aprendizaje automáticos y cuales han sido las razones.

En la Medicina, La utilización de algoritmos de Aprendizaje Automático, en los sistemas de diagnóstico, estos con la ayuda de la analítica, puede conllevar a mejoras importantes en la predicción de muchas enfermedades, para poder funcionar adecuadamente, el Machine Learning por lo general necesita enormes cantidades de datos, en tal sentido su funcionamiento y despliegue está íntimamente vinculado con otra nueva tecnología como el Big Data; afortunadamente, muchos sistemas de salud, han optado por la digitalización. por tanto, el historial clínico en general de los pacientes y sus tratamientos ya están en formato digital, de esta manera el sistema puede alimentarse de ellos para encontrar datos útiles y ejercer la analítica y medicina predictiva.

2) Indique algún escenario apropiado para la utilización de Machine Learning

Sugerencias de compras: Un sitio web bien puede sugerir determinados productos evaluando los que anteriormente alguien ya ha comprado, a partir de la comparativa de sus compras con las de otros compradores, el clasificador le podría sugerir la compra de productos que otros ya han comprado, si tú has comprado productos similares a los suyos.

Reconocimiento facial, esto tiene relación con el aprendizaje de tener algunos elementos del rostro previamente almacenados y haciendo que de alguna forma se registre o se entrene y se almacene en una base de datos, los rasgos identificativos de un rostro, para no solo detectar, sino también identificar a una persona.

3) Cuales serian las secuencias o los pasos para poder utilizar el aprendizaje automático

1 Colectar Datos: En cada problemática que se desea resolver, se debe investigar y obtener datos que se utilizarán para alimentar a tu máquina.

2 Preparación de los datos: Es importante realizar visualizaciones de nuestros datos, para revisar y editar si es que hubieren correlaciones entre las distintas características, separar y organizar; aquí también se puede

preprocesar datos realizando carpintería de datos, normalizando, eliminar duplicados y haciendo corrección de errores.

3 Elegir el modelo: Existen diferentes modelos que pueden elegirse de acuerdo al objetivo que se tenga: predicción, clasificación, regresión lineal, clustering, Deep Learning, bayesiano, etc. y podrá haber variantes si lo que se va a procesar son imágenes, sonido, texto, valores numéricos

4 Entrenar nuestra máquina: Se utiliza un set de datos de entrenamiento para ejecutarla y se debe ver una mejora incremental (para las predicciones). Recordar inicializar los "pesos" del modelo de forma aleatoria

5 Evaluación: Se debe comprobar que la máquina creada contra nuestro set de datos de Evaluación que contiene entradas que el modelo desconoce y en este caso hacer la verificación de la precisión del modelo ya entrenado; si es que la exactitud es menor o igual al 50%, es un modelo no útil; Si se alcanza un 90% o más se puede decir que existe buena confianza en los resultados que otorga el modelo.

6 Parameter Tuning (configuración de parámetros): Este "tuneo" es más un arte que una ciencia, y se irá mejorando a medida que se vaya experimentando, ya que suele haber muchos parámetros que se van ajustando y al combinarlos se pueden disparar todas las opciones buscadas.

7 Predicción o Inferencia: Ya listos para utilizar el modelo de Aprendizaje Automático, con nueva información y se da paso a la predicción e inferencia de resultados, "en la vida real" Imaginemos que si este paso fuera pasar de un simulador de un coche a aplicar nuestra máquina en un coche de verdad... es un gran momento y también será un emocionante desafío poner a prueba realmente todo nuestro trabajo de meses

4) Existe alguna librería o como se invoca esta tecnología con lenguajes de programación o algún dispositivo, explicar su experiencia

~~Matplotlib~~ es una la librería gráfica de Python, se puede usar para generar gráficos de calidad necesaria para publicarlas tanto en papel como digitalmente, esto es: series temporales, histogramas, espectros de potencia, diagramas de barras, diagramas de errores, etc.

~~Scikit-learn~~ es una librería de Python también, usada para Machine Learning y Análisis de Datos. Es basada en ~~Numpy~~, ~~SciPy~~ y ~~Matplotlib~~; de las ventajas principales se tiene, su facilidad de uso y la gran cantidad de técnicas de aprendizaje automático que permite implementar.