

CAMPUS FUENLABRADA

UNIVERSIDAD REY JUAN CARLOS

Titulo del Trabajo Fin de Máster

TRABAJO FIN DE MÁSTER

MÁSTER UNIVERSITARIO EN FORMACIÓN DEL
PROFESORADO DE ED.SECUNDARIA, BACHILLERATO,
FP E IDIOMAS

AUTOR: Abel de Andrés Gómez

TUTOR/ES: Nombre y Apellidos y

Nombre y Apellidos

AGRADECIMIENTOS

Agrademos a...

RESUMEN

Extensión máxima de una página

SUMMARY

Extensión máxima de una página

Índice

1.	INTRODUCCIÓN	1
2.	JUSTIFICACIÓN TEÓRICA	3
3.	PROPUESTA DE INTERVENCIÓN	7
4.	DISEÑO DE INVESTIGACIÓN	9
4.1.	Identificación de necesidades, fuentes de información y medios de acceso	10
4.1.1.	Identificación de necesidades de información	10
4.1.2.	Identificación de fuentes internas y externas de información . .	10
4.2.	Planificación de la realización de la vigilancia e inteligencia	11
4.3.	Busqueda y tratamiento de la información	11
4.3.1.	Proceso de Extracción, Extracción y Carga	11
4.3.2.	Análisis Descriptivo	12
4.3.3.	Machine Learning	12
4.4.	Distribución y Almacenamiento	15
5.	ANÁLISIS DE RESULTADOS	17
6.	CONCLUSIONES	19
7.	SOBRE LAS REFERENCIAS	21

Índice de figuras

1.	Proceso de la vigilancia e inteligencia	9
2.	Funcionamiento Árboles Decisión	14
3.	Teorema de Bayes	14
4.	Red Neuronal	15

Índice de cuadros

1. INTRODUCCIÓN

En los últimos años, gracias al gran desarrollo tecnológico que se ha vivido tanto a nivel de computo (mejorando la eficiencia y el uso de los recursos disponibles) como a nivel de transmisión de datos (mejorando las comunicaciones), ha permitido a las organizaciones el almacenamiento de una gran cantidad de información.

Para comprender mejor este gran volumen de información, es necesario utilizar métodos, técnicas, herramientas además de personas con conocimientos (formando todas esta un vínculo estrecho) que permita y ayude a explotar, investigar, predecir y obtener información relevante para tomar decisiones de forma adecuada.

La organización educativa no ha quedado ajena a estas necesidades de una mejor comprensión de los datos. En este sentido, una unidad de Educación Secundaria Obligatoria de la Consejería de Educación de la Comunidad de Madrid ha planteado un problema.

El problema con el que se enfrentan cada día es la planificación de grupos para el siguiente curso. Esta planificación es la base para poder decidir donde se escolariza cada alumno y como se va a repartir la plantilla del profesorado según sus especialidades. Conocer el número de grupos permite, por tanto, un óptimo reparto de la plantilla de docentes y de recursos. De esta forma, además, se evita la existencia de grupos sobrepoblados.

Desde esta unidad, han informado sobre aspectos sobre los que trabajan para poder realizar una predicción acerca del número de grupos para curso venidero.

Estos aspectos son:

1. Escolaridad del curso actual.

- Número de alumnos y grupos de un determinado centro.
- El número de alumnos por aula (también conocido como ratio).
- Matriculación de nuevos alumnos.
 - Principalmente alumnos que superan el nivel de 6º de primaria y pasan a 1º de ESO.

2. Bilingüismo del centro. Muchos alumnos optan por centros bilingües para su mejor formación, por lo que estos centros suelen tener más demanda de alumnos.

3. Posibilidad de creación de nuevas zonas urbanas cerca del centro.

4. Posibilidad de apertura o cierre de centros educativos. El cierre de por ejemplo, de un centro privado, provocara una mayor tasa de matriculación de los centros contiguos.

5. Porcentaje de aprobados. Los alumnos que están ya matriculados tienen prioridad sobre los nuevos alumnos, por lo tanto, si existe una alta tasa de suspensos, quedan pocas plazas de admisión de nueva matrícula.
6. El número y la aparición de nuevas enseñanzas. La oferta de nuevas enseñanzas atraerá a nuevos alumnos al centro, incrementando así el número de matriculaciones.

La unidad actualmente utiliza herramientas manuales para conseguir conocer el número de grupos, indicando que es un trabajo mecánico y con herramientas obsoletas, evitando la posibilidad de inclusión de nuevas variables o factores que impliquen nuevos resultados.

Propone dar una solución al problema actual mediante el uso de herramientas y métodos que automaticen dichas tareas y proponga, además, nuevas variables o factores que puedan influir en la toma de decisión.

2. JUSTIFICACIÓN TEÓRICA

En la actualidad existen numerosos informes acerca del uso de la ciencia de datos y sus técnicas en el ámbito educativo. Para la realización de este TFM se han analizado distintas publicaciones de la base de datos científica de ScienceDirect. Para realizar la búsqueda se han utilizado las siguientes palabras claves: educational, data y mining. Se debe recordar que el éxito de la búsqueda depende de estas palabras claves.

De la búsqueda anterior se han obtenido 160 artículos. Posteriormente se han seleccionado aquellos de los últimos 4 años (2016,2017,2018 y 2019). De esta forma obtenemos resultados actuales. Filtrando por fechas, hemos conseguido reducir los resultados a 73 artículos. Se ha realizado una observación sobre los artículos obtenidos y se ha comprobado la existencia de artículos que no resultaban útiles en esta investigación. Por tanto, se ha realizado otra búsqueda utilizando las claves anteriores y añadiendo la clave "prediction". Esta vez, se han obtenido 26 resultados. De todos los resultados obtenidos, se han seleccionado 15 artículos que se consideran útiles y que servirán de ayuda.

En el artículo de prensa de Fernandes [2], se muestra el uso de técnicas como los métodos de clasificación y el algoritmo predictivo de GBM (Gradient Boosting Model) con el objetivo de obtener aquellas variables en el entorno del alumno, que hace que este obtenga mejores o peores resultados escolares. Este estudio, además, tiene el objetivo de aportar información útil para los representantes políticos en el ámbito educativo, el consejo escolar y los profesores con el objetivo de que estos puedan realizar políticas públicas, materiales didácticos y trabajo social para beneficiar a los estudiantes.

Los datos escolares a estudiar proceden de alumnos de colegios de un Distrito Federal de Brasil durante el 2015 y el 2016. Estos datos se han obtenido a partir de la base de datos de iEducar que contiene atributos relacionados con cada alumno.

Algunas de las variables que se estudian en el artículo anterior pertenecen concretamente al ámbito personal, social y geográfico del alumno. Estas variables son:

1. El barrio del alumno.
2. El centro educativo.
3. La edad del alumno.
4. Los ingresos del alumno.
5. Los alumnos con necesidades especiales.
6. El género.
7. El entorno en el aula.

En esta investigación se utiliza la metodología CRISP-DM (del inglés Cross Industry Standard Process for Data Mining) que es una metodología frecuente en el desarrollo de proyectos de Data Mining.

Para realizar la parte de prediccion, utiliza las variables anteriormente comentadas, incluidas en dos conjuntos de datos. En el primer conjunto de datos (DS-I), se almacenan los datos obtenidos antes de comenzar el comienzo del año escolar. El segundo conjunto de datos incluye las variables del primer conjunto de datos y alguna nueva que se ha obtenido despues del segundo mes del año escolar. Siendo algunas de estas variables nuevas las asignaturas, las notas y las ausencias. Estas dos ultimas variables son las que mayor importancia tienen en la revelacion de los resultados academicos finales.

El primer conjunto de datos (DS-I) se usa para entrenar el modelo de clasificacion I (CM-I), que identifica la probabilidad que tiene un alumno de suspender teniendo en cuenta los datos del comienzo de curso. El segundo conjunto de datos (DS-II) se usa para entrenar el modelo de clasificacion II, que tambien identifica la probabilidad que tiene un alumno de suspender teniendo en cuenta los datos del comienzo de curso e incluyendo las nuevas variables. Una vez que se han entrenado los modelos, se ha utilizado la matriz de confusion para obtener la bondad o efectividad del modelo respecto al conjunto de datos. Los datos obtenidos han mostrado que las variables de “vecindario”, “colegio”, “ciudad”, “edad” son factores relevantes que afectan a los resultados acadameicos de los alumnos.

Como conclusion, se indica en esta investigacion que el entorno social y sus variables tienen una influencia directa en el proceso de enseñar-aprender. Esta investigacion puede aportar informacion a los profesionales que busquen herramientas o metodos para mejorar los resultados escolares de los alumnos.

Por otro lado, en el articulo de prensa de R.Asif et al. [1] se citan otras investigaciones realizadas, donde tambien se utilizan variables sociales como la edad, sexo, nacionalidad, estado civil, desplazamiento (si el alumno vive fuera del distrito), necesidades especiales, tipo de admision, situacion laboral, situacion economica, etc.

En el articulo de R.Asif et al. [1], se analiza el rendimiento de los alumnos matriculados en el 4 año del grado universtario de Tecnologia Informática. El objetivo es, nuevamente, obtener informacion sobre el rendimiento de estudiantes para que las personas interesadas (directores y docentes) puedan mejorar el programa educativo. Los enfoques para lograr este objetivo son los siguientes:

1. En primer lugar se generan clasificadores para predecir el rendimiento de los estudiantes al final del curso academico tan pronto como sea posible. Estos clasificadores se toman la calificaciones de admision y las calificaciones finales del primer y segundo año. No se consideran caracterisitcas socioeconomicas o demograficas.
2. En segundo lugar, utilizando estos clasificadores, el objetivo es utilizar cursos que puedan servir como indicadores efectivos del desempeño de los estudiantes. De esta forma se puede ayudar o estimular a los alumnos en riesgo.

3. Por ultimo, se va a investigar como el rendimiento academico progresa sobre el cuarto año del grado. Para ello, se va a utilizar tecnicas de *clustering* y se van a dividir a los alumnos en grupos, donde los alumnos de un mismo grupo van a tener la misma progresion en el rendimiento. De esta forma, se van a agrupar los alumnos que hayan tenido bajas calificaciones a lo largo de sus estudios y aquellos que han tenido altas calificaciones a lo largo de sus estudios. La clave es obtener y comprender los indicadores propuestos en el segundo paso.

Los datos utilizados en esta investigacion proceden de las calificaciones del cuarto año del grado de ingenieria de Tecnologia Informática de una universidad de Pakistan. Se van a tomar 210 alumnos que se han matriculado en los cursos de 2007-2008 y 2008-2009. Los datos contienen variables relacionadas con las calificaciones de pre-admision de los alumnos y de las calificaciones de estos en los siguientes 4 años del programa de grado. Por tanto, para lograr los objetivos establecidos, R.Asif et al. [1], va a utilizar los arboles de decision y clustering como tecnicas de mineria de datos.

3. PROPUESTA DE INTERVENCIÓN

Con el objetivo de resolver el problema comentado en los apartados anteriores, se plantea el uso de la ciencia de datos como proceso para descubrir relaciones entre los datos, que sean significativas. Además, se van a buscar patrones y tendencias en los datos que ayuden a la toma de decisiones.

En primer lugar, se debe tener en cuenta que la ciencia de datos aúna métodos y tecnologías que provienen del campo de las matemáticas, la estadística y la informática entre las que se pueden encontrar el análisis descriptivo o exploratorio, el aprendizaje automático (“machine learning”), el aprendizaje profundo (“Deep learning”), etc. [4]. En esta propuesta de intervención, se va a centrar en el análisis descriptivo y el aprendizaje automático.

El análisis descriptivo, como ya se ha comentado, va a ser útil para observar características de los propios datos. Entre estas características se va a poder observar cuales son las variables que más convienen al estudio por su importancia, utilizando técnicas como el análisis principal de componentes. Se puede observar también la correlación entre las variables, sobretodo.

El aprendizaje automático, se divide en dos áreas: el aprendizaje supervisado y el no supervisado.

- El aprendizaje supervisado se basa en algoritmos que intentan encontrar una función, que, dadas las entradas, asigne unas salidas adecuadas. Estos algoritmos se entrenan mediante datos históricos y de esta forma aprende a asignar salidas adecuadas en función de dichas entradas, dicho de otra forma, predice el valor de salida. A su vez, el aprendizaje supervisado se divide en regresión (si la salida es de tipo numérico) y clasificación (si la salida es del tipo categórico). [7]
- El aprendizaje no supervisado se utiliza en datos en los que existen variables de entrada, pero no existen variables de salida para dichas variables de entrada. Por consiguiente, solo se puede describir la estructura de los datos, para intentar conseguir algún tipo de estructura u organización que simplifique el análisis.[7]

4. DISEÑO DE INVESTIGACIÓN

Uno de los pilares básicos en el diseño de una investigación es indicar el camino que se va a seguir en esta. Es importante establecer que estándar o norma se va a seguir en el desarrollo de un proyecto o una investigación. En esta investigación se va a utilizar la norma UNE 166006:2018 Gestión de la I+D+I: Sistemas de vigilancia e inteligencia. Esta norma está alineada con la norma UNE-EN ISO 9001 Sistema de Gestión de Calidad.

La norma UNE 166006:2018 tiene como objeto facilitar la formación y estructuración del proceso de recogida, análisis y comunicación de la información sobre el entorno de una organización. No solo muestra un proceso, sino que también establece roles, responsabilidades y políticas.

La UNE 166006:2018 establece un proceso generico para satisfacer los objetivos deseados y contemplar la realizacion de la vigilancia e inteligencia. Este proceso se divide en distintas etapas basicas. En la imagen 1 se puede observar cada etapa.

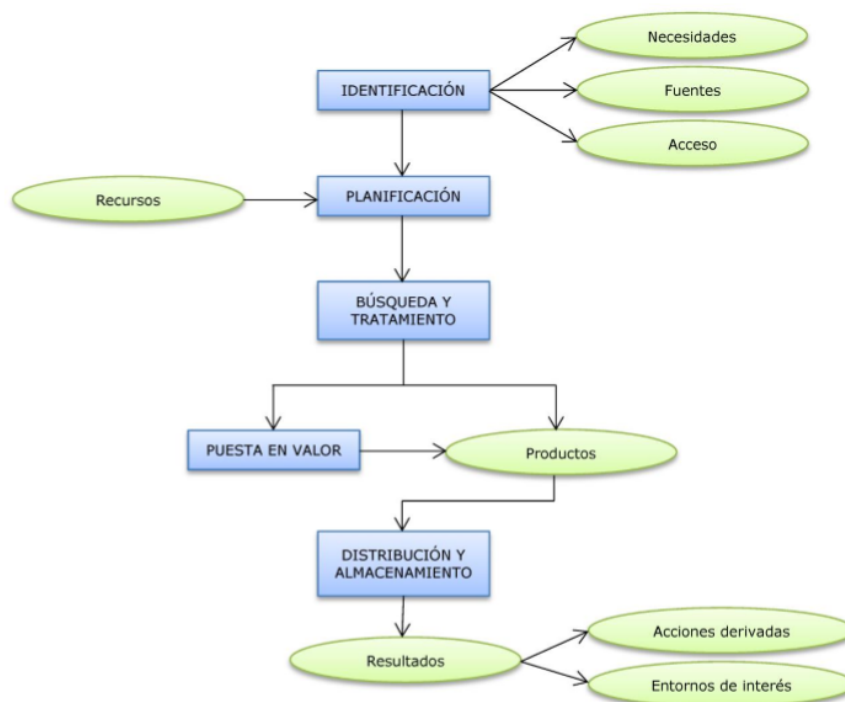


Fig. 1: Proceso de la vigilancia e inteligencia

En los proximos puntos se va a describir las actividades que se van a realizar en este proceso.

4.1. Identificación de necesidades, fuentes de información y medios de acceso

4.1.1. Identificación de necesidades de información

Para realizar la identificación de las necesidades de información se va a partir de varios factores como son:

- las demandas esperadas o manifestadas por (en este caso) una unidad de la consejería de educación.
- el análisis, la evolución de productos, procesos, materiales y tecnologías en el ámbito de la minería de datos educativos.

4.1.2. Identificación de fuentes internas y externas de información

Teniendo en cuenta las principales necesidades de información, se debe identificar las fuentes de información y recursos disponibles ya sean internos o externos a la organización. En este caso, se van a utilizar las siguientes fuentes:

- Fundamentalmente se va a utilizar documentos y recursos internos de la organización como van a ser:
 - Repositorios documentales.
 - Carpetas locales.
 - Bases de datos.
 - Etc.
- Fuentes documentales a las que tiene acceso la organización, ya sea en soporte físico (revistas, catálogos, etc.) como en soporte electrónico. Además se utilizarán recursos de información en Internet (portales, noticias, redes sociales, foros, etc).
- Personas con conocimientos o experiencias relacionadas con las necesidades de información. En este aspecto se van a realizar distintas reuniones con las personas encargadas de esta unidad de la consejería de educación.
- Documentación técnica como reglamentaciones, especificaciones, propiedad industrial e intelectual o normas.
- Resultados de análisis existentes sobre las tendencias de futuro preferentemente en el ámbito educativo.

4.2. Planificación de la realización de la vigilancia e inteligencia

Para realizar la planificación del proceso, se debe tener en cuenta si se requiere el uso de nuevas áreas desconocidas que puedan llevar a la búsqueda y a la investigación, o si se va a realizar un seguimiento sistemático de áreas que ya están identificadas. En el caso de esta investigación, se va a utilizar métodos y técnicas ya identificadas, sin embargo como se puede esperar, los objetivos son distintos a otras investigaciones.

4.3. Búsqueda y tratamiento de la información

La información fundamentalmente se encontrará en bases de datos internas, aunque se va acceder a bases de datos externas en caso de necesidad para complementar la información. En este aspecto, se debe recurrir a la ayuda de personas con conocimientos sobre el estado de las bases de datos. Como cualquier organización, la consejería de educación maneja grandes volúmenes de datos, por tanto, se debe tener conocimiento sobre donde se puede encontrar la información que satisfaga con las necesidades. El desconocimiento del estado de las bases de datos conlleva la inversión de una gran cantidad de tiempo en la búsqueda de los datos relevantes.

Una vez que se tienen los datos, muchas veces es necesario realizar un tratamiento de estos, que consiste en una limpieza y una normalización de los mismos. Muchas veces este tratamiento conlleva la conversión de datos, como por ejemplo fechas, corrección de datos, etc.

4.3.1. Proceso de Extracción, Transformación y Carga

Para realizar este tratamiento de datos se utilizará la técnica conocida como ETL (extracción, transformación y carga) que consiste básicamente en obtener los datos de la fuente de origen (bases de datos, ficheros Excel, ficheros JSON, etc), seleccionar aquellos datos que convengan al estudio, transformarlos según las necesidades que se tenga y depurarlos (evitando así datos erróneos). (Prakash, 2017) (Guillermo Matos, 2006) (Sharma, 2014). Para realizar este tratamiento, se ha va a utilizar Pentaho BI, que es un conjunto de programas libres para realizar entre otras muchas actividades, las técnicas de ETL. Concretamente, se ha utilizado la herramienta Spoon para desarrollar esta técnica. Una vez que se tienen los datos limpios y estructurados, se pueden realizar dos operaciones:

1. En primer lugar, se pueden almacenar dichos datos en una base de datos y seguir utilizando Pentaho BI para poder crear cuadros de mandos e informes.
2. En segundo lugar, se puede almacenar la información en un texto plano para poder trabajar con herramientas de análisis descriptivo y predictivo. Estos análisis se van a realizar a través del entorno y lenguaje de programación R, que es una referencia en el ámbito de la estadística.

4.3.2. Análisis Descriptivo

El análisis predictivo (también conocido como estadísticas predictivas) se encarga de resumir los datos en bruto para que puedan ser interpretados. Estos análisis son útiles ya que permiten aprender sobre comportamientos o patrones pasados e entender cómo pueden influir en los resultados futuros. En este tipo de análisis se van a utilizar tanto métodos gráficos como medidas resumen.

En primer lugar, se debe estudiar el tipo de datos de cada variable a estudiar, se debe clasificar las variables según sean categóricas (dicotómicas o polinómicas) o numéricas (discretos o continuos). El tipo de datos permite decidir qué tipo de análisis estadístico utilizar. Una vez que se tienen claro el tipo de datos utilizados, se van a utilizar los principales estadísticos como la media, la mediana, las desviaciones típicas, etc. Posteriormente se va a utilizar la matriz de varianzas y covarianzas, que indicaran la variabilidad de los datos y la información sobre las posibles relaciones lineales entre las variables.

Por otro lado, se va a estudiar la correlación de las variables mediante la matriz de correlación. Esta matriz contendrá los coeficientes de correlación.[5]. La matriz de correlación, se utilizará fundamentalmente por pares entre las variables y la variable a predecir.

También se va a estudiar la matriz de correlaciones parciales, que estudia la correlación entre pares de variables eliminando el efecto de las restantes.[5]

Los datos categóricos se van a representar en tablas de frecuencias, gráficos de barras y gráficos de sectores. Los datos numéricos se van a representar mediante histogramas, boxplot y diagramas QQ-Plot o Grafico Cuantil-Cuantil. [6]

Mediante el boxplot se puede observar aspectos como la posición, dispersión, asimetría, longitud de colas y los datos anómalos (outliers). El QQ-plot se va a utilizar para evaluar la cercanía de los datos a una distribución. [6]

Por otro lado, se va a complementar el análisis descriptivo mediante el aprendizaje no supervisado, donde también se extraerán otras características de los datos.

4.3.3. Machine Learning

Aprendizaje No Supervisado

1. Algoritmos de Clustering. El objetivo de estos algoritmos consiste en investigar si los datos pueden ser organizados en grupos o *clusters* que posean características similares. Los metodos de clustering tienen la caracteristica comun que para poder llevar a cabo las agrupaciones necesitan definir y cuantificar la similitud entre las observaciones. Por ejemplo la distancia euclidea, la distancia de Manhattan, la correlacion, el indice de Jaccard, etc. Para realizar este analisis, se va a utilizar el algoritmo de K-Means.

2. Análisis de Componentes Principales
3. Descomposición en valores singulares
4. Analisis de componentes independientes
5. Stepwise Regression

Aprendizaje Supervisado

Una vez terminado el análisis descriptivo, se va a realizar un análisis predictivo. Se debe tener en cuenta, que, dentro de la ciencia de datos, existen técnicas de aprendizaje automáticas, cuyo objetivo es la construcción de un sistema que sea capaz de aprender a resolver problemas sin la intervención de un humano. [4].

El aprendizaje supervisado consiste en la búsqueda de patrones en datos históricos relacionando todas las variables con una especial (conocida como variable objetivo). Los algoritmos que se utilizan en el aprendizaje supervisado se encarga de buscar patrones en los datos. A este proceso se conoce como entrenamiento de los datos. Una vez que se tienen los patrones, se aplican a los datos de prueba. Los datos de entrenamiento suelen ser una selección aleatoria y única de los datos históricos de un 70 % del total. Los datos de prueba son el restante 30 %. [3] Algunos de los algoritmos que se van a utilizar son:

1. Arboles de decisión

Se basa en el descubrimiento de patrones a partir de ejemplos. Un árbol de decisión está formado por un conjunto de nodos (de decisión) y de hojas (nodos-respuesta).

Los nodos están asociados a los atributos y tiene varias ramas que salen de él (dependiendo de los valores que tomen la variable asociada). Estos nodos pueden asemejarse a preguntas que, dependiendo de la respuesta que conlleve, se tomara un flujo en las ramas salientes.

Los nodos respuesta están asociados a la clasificación que se desea proporcionar, devolviendo así la decisión del árbol con respecto al ejemplo de entrada utilizado.



Fig. 2: Funcionamiento Árboles Decisión

2. Clasificación de Naïve Bayes.

Es un algoritmo que se basa en la técnica de clasificación utilizando el teorema de Bayes.

El algoritmo es capaz de agrupar un registro mediante las características de este. Para ello aplica probabilidades condicionales de las características para determinar a qué categoría pertenece.

Por ejemplo, una fruta puede considerarse una manzana si es de color rojo, redonda y tiene un determinado peso.

$$P(c|x) = \frac{P(x|c)P(c)}{P(x)}$$

Likelihood $\rightarrow P(x|c)$
 Class Prior Probability $\rightarrow P(c)$
 Posterior Probability $\leftarrow P(c|x)$
 Predictor Prior Probability $\leftarrow P(x)$

$$P(c|X) = P(x_1|c) \times P(x_2|c) \times \dots \times P(x_n|c) \times P(c)$$

Fig. 3: Teorema de Bayes

3. Regresión Logística

Es un algoritmo de regresión que se utiliza para predecir el resultado de una variable categórica en función de las variables independientes o predictoras. Para predecir el resultado, se establecen pesos en función de la puntuación dada a cada variable independiente.

4. Redes Neuronales

Las redes neuronales son un algoritmo de inteligencia artificial que se inspira en los mecanismos presentes en la naturaleza. Las neuronas envían señales eléctricas de manera fuerte o débil a otras neuronas. La combinación de todas las conexiones entre neuronas es lo que genera el conocimiento. Estas señales se envían cuando existen unos estímulos (inputs) externos a través de los sentidos. A lo largo de la vida, las neuronas aprenden que deben hacer a partir de dichos estímulos y por lo tanto, los seres vivos aprenden a actuar

ante distintas señales y situaciones. El funcionamiento de las redes neuronales en la inteligencia artificial es similar.

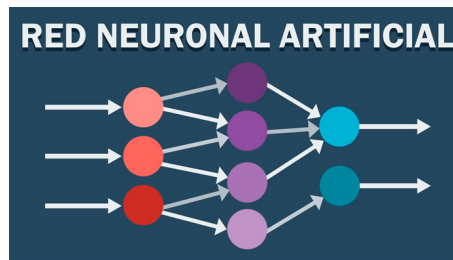


Fig. 4: Red Neuronal

Como se puede observar en la figura 4, la primera fila (con neuronas de color rojo), se conocen como nodos de entrada y son aquellos que se encargan de recoger la información. Los nodos en la gama azul son los que se conocen como nodos de salida. Los nodos situados en el medio son aquellos que se encargan de hacer el aprendizaje, y se conocen como nodos ocultos.

En primer lugar, se obtiene la información a partir de los nodos de entrada, una vez que se tiene la información, se envía a las capas ocultas, que se activan o no dependiendo del aprendizaje previo. Los nodos ocultos se activan dependiendo de una serie del resultado de unas operaciones matemáticas. Si los nodos se activan, entonces enviarán la información a la siguiente capa.

5. Potenciación del gradiente Es un método que se encarga de construir modelos de manera secuencial, cada uno de ellos se centra en los datos que el anterior modelo clasificaba mal.
6. Random Forest
7. Combinación de modelos

4.4. Distribución y Almacenamiento

Respecto a la distribución de la información, esta no podrá salir de la consejería de educación. Aunque se trate de datos anonimizados y agregados, se trata de datos de carácter sensible y no pueden ser distribuidos. Por otro lado, dichos datos se almacenarán en un gestor de bases de datos MySQL. Este gestor se encontrará en un servidor de la Consejería de Educación. Solo se va a poder acceder a dicho servidor desde la propia sede.

5. ANÁLISIS DE RESULTADOS

6. CONCLUSIONES

7. SOBRE LAS REFERENCIAS

La bibliografía o referencias deben aparecer siempre al final de la tesis, incluso en aquellos casos donde se hayan utilizado notas finales. La bibliografía debe incluir los materiales utilizados, incluida la edición, para que la cita pueda ser fácilmente verificada.

Citar dentro del texto:

Las fuentes consultadas se describen brevemente dentro del texto y estas citas cortas se amplían en una lista de referencias final, en la que se ofrece la información bibliográfica completa.

La cita dentro del texto es una referencia corta que permite identificar la publicación de donde se ha extraído una frase o parafraseado una idea, e indica la localización precisa dentro de la publicación fuente. Esta cita informa del apellido del autor, la fecha de publicación y la página (o páginas) y se redacta de la forma que puede verse a través de los siguientes ejemplos:

Cuando se citan las palabras exactas del autor deben presentarse entre comillas e indicarse, tras el apellido del autor y, entre paréntesis, la fecha de publicación de la obra citada, seguida de la/s página/s.

Si lo que se reproduce es la idea de un autor (no sus palabras exactas) no se pondrán comillas y se indicará, entre paréntesis, el apellido del autor seguido de la fecha de publicación de la obra a la que se refiere.

No se puede eliminar una parte del texto citado sin señalarse; debe indicarse siempre con puntos suspensivos entre corchetes [...]

Ejemplos de cómo citar una referencia en el texto son los siguientes [?] o [?, ?, ?].

Cómo ordenar las referencias:

1. Las referencias bibliográficas deben presentarse ordenadas alfabéticamente por el apellido del autor, o del primer autor en caso de que sean varios.
2. Si un autor tiene varias obras se ordenarán por orden de aparición.
3. Si de un mismo autor existen varias referencias de un mismo año se especificarán los años seguidos de una letra minúscula y se ordenarán alfabéticamente.
4. Si son trabajos de un autor en colaboración con otros autores, el orden vendrá indicado por el apellido del segundo autor, independientemente del año de publicación. Las publicaciones individuales se colocan antes de las obras en colaboración.

Cómo citar un artículo de revista

Un artículo de revista, siguiendo las normas de la APA, se cita de acuerdo con el siguiente esquema general: Apellido(s), Iniciales del nombre o nombres. (Año de publicación). Título del artículo. Título de la revista en cursiva, volumen de la

revista (numero del fascaculo entre parantesis), primera pagina- ultima pagina del artaculo.

Camo citar una monografaa/libro

Las monografaas, siguiendo las normas de la APA, se citan de acuerdo con el siguiente esquema general: Apellido(s), Iniciales del nombre. (Aao de publicacion). *Tatulo del libro* en cursiva. Lugar de publicacion: Editorial. Opcionalmente podremos poner la mención de edición, que ira entre parantesis a continuación del título; y, si fuera el caso el volumen que ira en cursiva.

Camo citar un capitulo de un libro

Los capítulos de los libros se citan de acuerdo con el siguiente esquema general: Apellido(s), Iniciales del nombre o nombres. (Aao). *Tatulo del capitulo*. En A. A. Apellido(s) Editor A, B. B. Apellido(s) Editor B, y C. Apellido(s) Editor C (Eds. o Comps. etc.), *Tatulo del libro* en cursiva (pp. xxx-xxx). Lugar de publicacion: Editorial.

Camo citar un acta de un congreso

Apellido(s), Iniciales del nombre o nombres. (Aao). *Tatulo del trabajo*. En A. A. Apellido(s) Editor A, B. B. Apellido(s) Editor B, y C. Apellido(s) Editor C (Eds. o Comps. etc.), *Nombre de los proceedings* en cursiva (pp. xxx-xxx). Lugar de publicacion: Editorial.

Camo citar tesis doctorales, trabajos fin de master o proyectos fin de carrera

Apellido(s), Nombre. (Aao). *Tatulo de la obra* en cursiva. (Tesis doctoral). Institucion a academica en la que se presenta. Lugar.

Camo citar un recurso de Internet

Los recursos disponibles en Internet pueden presentar una tipología muy variada: revistas, monografaas, portales, bases de datos... Por ello, es muy difícil dar una pauta general que sirva para cualquier tipo de recurso. Como mínimo una referencia de Internet debe tener los siguientes datos:

1. *Tatulo y autores del documento.*
2. *Fecha en que se consulta el documento.*
3. *Dirección (URL a uniform resource locator)*

Veamos, a través de distintos ejemplos, como se citan específicamente algunos tipos de recursos electrónicos.

Monografaas: Se emplea la misma forma de cita que para las monografaas en versión impresa. Debe agregar la URL y la fecha en que se consulta el documento

Articulos de revistas: Se emplea la misma forma de cita que para los artículos de revista en versión impresa. Debe agregar la URL y la fecha en que se consulta el documento.

Articulos de revistas electronicas que se encuentran en una base de datos: Se emplea la misma forma de cita que para los articulos de revista en version impresa, pero debe aadirse el nombre de la base datos, la fecha en que se consulta el documento.

ANEXOS

BIBLIOGRAFÍA

Referencias

- [1] Asif, R., Merceron, A., Abbas Ali, S., & Ghani Haider, N. (2017, 22 mayo). Analyzing undergraduate students' performance using educational data mining. Elsevier, Computers & Education(113), 177–194.
- [2] Fernandes, E., Holanda, M., Victorino, M., Borges, V., Carvalho, R., & Van Erven, G. (2018, 7 febrero). Educational data mining: Predictive analysis of academic performance of public school students in the capital of Brazil. Journal of Business Research, pp. 1–9.
- [3] Manguart, A. (2017, 13 junio). Conceptos básicos del aprendizaje supervisado (para personas no técnicas) [Publicación en un blog]. Recuperado 15 enero, 2019, de <https://medium.com/@manguart/machine-learning-conceptos-basicos-del-aprendizaje-supervisado-para-personas-no-tecnicas-142bbb222140>
- [4] Marín, J. L. (2018, 5 abril). *Ciencia de datos, machine learning y deep learning* (Comunicado de prensa). Recuperado 17 enero, 2019, de <https://datos.gob.es/es/noticia/ciencia-de-datos-machine-learning-y-deep-learning>
- [5] Marín, J. M. (s.f.). Estadística Descriptiva Univariante [Tema Universitario]. Recuperado 17 enero, 2019, de <http://halweb.uc3m.es/esp/Personal/personas/jmmarin/esp/AMult/tema2am.pdf>
- [6] Orellana, L. (2001). Introducción. In L. Orellana (Ed.), Estadística Descriptiva (pp. 1–64). Recuperado de http://www.dm.uba.ar/materias/estadistica_Q/2011/1/modulo%20descriptiva.pdf
- [7] Recuero, P. (2017, 16 noviembre). Los 2 tipos de aprendizaje en Machine Learning: supervisado y no supervisado [Publicación en un blog]. Recuperado 17 enero, 2019, de <https://data-speaks.luca-d3.com/2017/11/que-algoritmo-elegir-en-ml-aprendizaje.html>