



DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

**“SISTEMA WEB PARA APOYAR EN LA LECTOESCRITURA A NIÑOS
CON DIFICULTADES ESPECÍFICAS DE APRENDIZAJE MEDIANTE
UN ALGORITMO DE APRENDIZAJE AUTOMÁTICO”**

TESIS

PRESENTADA COMO REQUISITO PARCIAL PARA OBTENER EL GRADO
DE:

MAESTRO EN CIENCIAS DE LA COMPUTACIÓN

ING. VICTOR MANUEL MORENO GARCIA

DIRECTOR:
M.C. CÉSAR ENRIQUE ROSE GÓMEZ

HERMOSILLO SONORA, MÉXICO

25 DE AGOSTO DEL 2021





EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Hermosillo
División de Estudios de Posgrado e Investigación

SECCIÓN: DIV. EST. POS. E INV.
No. OFICIO: DEPI/188/21
ASUNTO: AUTORIZACIÓN DE IMPRESIÓN
DE TESIS.

16 de agosto de 2021

**C. VICTOR MANUEL MORENO GARCÍA,
PRESENTE.**

Por este conducto, y en virtud de haber concluido la revisión del trabajo de tesis que lleva por nombre **"Sistema Web para Apoyar en la Lectoescritura a Niños con Dificultades Específicas de Aprendizaje Mediante un Algoritmo de Aprendizaje Automático"**; que presenta para el examen de grado de la MAESTRÍA EN CIENCIAS DE LA COMPUTACIÓN, y habiéndola encontrado satisfactoria, nos permitimos comunicarle que se autoriza la impresión del mismo a efecto de que proceda el trámite de obtención de grado.

Deseándole éxito en su vida profesional, quedo de usted.


ATENTAMENTE



M.C. CÉSAR ENRIQUE ROSE GÓMEZ
DIRECTOR


M.C. MARÍA TRINIDAD SERNA ENCINAS
SECRETARIA



S.E.P.
INSTITUTO TECNOLÓGICO
DE HERMOSILLO
DIVISIÓN DE ESTUDIOS
DE POSGRADO


DR. OSCAR MARIO RODRÍGUEZ ELÍAS
VOCAL


M.C. ROSA RENÉ SÁNCHEZ FERMÍN
JEFA DE LA DIVISIÓN DE ESTUDIOS DE POSGRADO E INVESTIGACIÓN

RISF/eme*



Av. Tecnológico S/N Col. El Sahuaro C.P. 83770 Hermosillo, Sonora
Tel. 01 (662) 260 65 00, ext. 136, e-mail: depi_hermosillo@tecnm.mx
tecnm.mx | ith.mx





EDUCACIÓN
SECRETARÍA DE EDUCACIÓN PÚBLICA



TECNOLÓGICO
NACIONAL DE MÉXICO

Instituto Tecnológico de Hermosillo
División de Estudios de Posgrado e Investigación

CARTA CESIÓN DE DERECHOS

En la ciudad de Hermosillo Sonora a el día 25 de Agosto del año 2021 el que suscribe C. Victor Manuel Moreno Garcia, alumno de la maestria en Ciencias de la Computación adscrito a la División de Estudios de Posgrado e Investigación, manifiesta que es autor intelectual del presente trabajo de Tesis titulado Sistema Web para apoyar a la lectoescritura a niños con dificultades especificas de aprendizaje mediante un algoritmo de aprendizaje automático bajo la dirección de César Enrique Rose Gómez y ceden los derechos del mismo al Tecnológico Nacional de México/Instituto Tecnológico de Hermosillo, para su difusión, con fines académicos y de investigación.

Los usuarios de la información no deben de reproducir el contenido textual, graficas, tablas o datos contenidos sin el permiso expreso del autor y del director del trabajo. Este puede ser obtenido a la dirección de correo electrónico siguiente: vmoreno85@hotmail.com. Una vez otorgado el permiso se deberá expresar el agradecimiento correspondiente y citar la fuente del mismo.

A T E N T A M E N T E

Victor Manuel Moreno Garcia



Av. Tecnológico S/N Col. El Sahuaro C.P. 83170 Hermosillo,
Sonora
Tel. 01 (662) 260 65 00, ext. 136, e-mail:
depi_hermosillo@tecnm.mx



Agradecimientos

A mis directores de tesis y maestros, por guiarme con su conocimiento y motivarme a nunca darme por vencido ante los nuevos retos, y por las oportunidades de aprender algo nuevo cada día.

A María Victoria, mi pequeña inspiración para iniciar este camino y a María Elena, mi pequeña inspiración para concluirlo...

A CONACyT por el apoyo brindado a través de la beca número 716906.

Resumen

Actualmente, los docentes de instituciones educativas ofrecen a sus alumnos diferentes actividades de apoyo a quienes padecen algún tipo de dificultad específica de aprendizaje, y aunque las actividades son efectivas, no siempre suelen generar el impacto adecuado en el alumno ya que suelen ser actividades extras para el docente y el alumno no tiene el apoyo suficiente para sobrellevarlas y avanzar de manera normal en su proceso de aprendizaje.

La presente tesis tiene como objetivo realizar una investigación y recopilación de información para proponer un sistema web que implementando un algoritmo de machine learning pueda recomendar una serie determinada de ejercicios de reforzamiento a los alumnos durante sus primeros tres años de educación básica, permitiéndoles resolver y repetir ejercicios acordes a sus dificultades de aprendizaje.

A lo largo del documento, se analizan una serie de diferentes algoritmos de inteligencia artificial para seleccionar el que más se adecue a las necesidades del proyecto, basándose principalmente en la precisión que pueden ofrecer al recibir el entrenamiento utilizando un set de datos de alrededor de 5000 registros de un formato alfabetización utilizado por la Secretaría de Educación Pública en la ciudad de Hermosillo, Sonora.

Finalmente, la implementación del algoritmo se da mediante el uso de una API desarrollada en Python que es consumida por un sistema web en PHP, diagnosticando alumnos en cuatro diferentes categorías según el vector de características registrado por el docente.

Abstract

Nowadays, teachers offer their students different support activities for those who suffer some type of specific learning disability, although the activities are effective, they are not always successful on the students since they are usually extra activities for both of them, and the learner do not have enough support to complete them and continue with their learning process.

The objective of this thesis is to carry out an investigation and compilation of information, to propose a web system that, by implementing a machine learning algorithm, can recommend certain series of reinforcement exercises to students during their first three years of basic education, allowing them to solve and repeat exercises according to their learning difficulties.

Throughout the document, a series of different artificial intelligence algorithms are analyzed to select the one that best suits the needs of the project, mainly based on their precision when receiving training using a set of data of around 5000 records of a literacy format used by the Secretaria de Educación Pública in the city of Hermosillo, Sonora.

Finally, the implementation of the algorithm is given through the use of an API developed in Python that is consumed by a web system in PHP, diagnosing students in four different categories according to the vector of characteristics registered by the teacher.

Índice

Capítulo 1	12
Introducción	12
1.1 Antecedentes.....	14
1.2 Definición del problema	16
1.3 Objetivos.....	17
1.4 Justificación	17
1.5 Limitaciones y alcances	18
1.6 Metodología.....	19
1.7 Organización de la tesis	20
Capítulo 2	22
Estado del arte	22
2.1 Dificultades específicas del aprendizaje.....	22
2.1.1 Dislexia	23
2.1.2 Disgrafía.....	33
2.1.3 Discalculia.....	37
2.1.4 Diagnóstico y tratamiento	40
2.2 Inteligencia artificial	43
2.2.1 Machine learning	44
2.3 Sistemas web	56
2.3.1 Elementos de un sistema web	57
2.3.2 Definición de cliente-servidor.....	57
2.3.3 Definición de Front End y Back End	58
2.4 Trabajos relacionados	59
2.4.1 Deslixate	60
2.4.2 Dytective	61
Capítulo 3	64
Análisis y diseño	64
3.1 Análisis del sistema	65
3.1.1 Arquitectura del sistema	66

3.1.2 Diagramas de contexto nivel 0.....	67
3.1.3 Diagramas de contexto nivel 1.....	67
3.1.4 Diagramas de casos de uso	68
3.2 Diseño del sistema.....	74
3.2.1 Diagramas conceptual de clases.....	74
3.2.2 Diagrama de secuencia	75
3.2.3 Diagramas de componentes y despliegue	77
3.2.4 Diagrama de actividades	78
3.2.5 Esquema de la base de datos	80
3.3 Proceso de selección de modelo de machine learning.....	81
Capítulo 4	83
Análisis de datos e implementación del algoritmo de clasificación.....	83
4.1 Preparación del set de datos.....	83
4.1.1 Recopilación de la información	83
4.1.2 Selección del vector de características.....	86
4.1.3 Proceso de normalización de datos	87
4.2 Diseño y desarrollo del algoritmo	88
4.2.1 Selección y entrenamiento de algoritmos de clasificación	88
4.2.2 Primera experimentación de los algoritmos.....	90
4.2.3 Resultados de primera experimentación de los algoritmos.....	91
4.2.4 Segunda experimentación de los algoritmos.....	92
4.2.5 Resultados de segunda experimentación de los algoritmos	93
4.2.6 Tercera experimentación de los algoritmos	93
4.2.7 Resultados de tercera experimentación de los algoritmos	94
4.2.8 Cuarta experimentación de los algoritmos.....	95
4.2.9 Resultados de cuarta experimentación de los algoritmos	96
4.4 Migración del algoritmo de machine learning.....	99
4.5 Creación de la base de datos	99
4.6 Configuración del entorno de trabajo	100
4.7 Instalación y configuración del Framework Laravel.....	100
4.8 Programación de usuarios y CRUD.....	101

Capítulo 5	104
Análisis de resultados	104
5.1 Resultados de las experimentaciones	104
5.2 Análisis de la matriz de confusión.....	105
Capítulo 6	107
Conclusiones y trabajo a futuro.....	107
6.1 Conclusiones.....	107
6.2 Trabajos a futuro.....	108
Referencias	110

Índice de tablas

Tabla 2.1 Clasificaciones de la dislexia propuestas por Josep Artigas	27
Tabla 2.2 Características diferenciales entre los distintos tipos de disgrafía	36
Tabla 3.1 Caso de uso: iniciar sesión	69
Tabla 3.2 Caso de uso: ejercicios	70
Tabla 3.3 Caso e uso: administrar grupos	71
Tabla 3.4 Caso e uso: administrar alumnos	72
Tabla 3.5 Caso de uso: reportes	73
Tabla 3.6 Caso de uso: módulo machine learning	74
Tabla 4.1 Resultados de primera experimentación	91
Tabla 4.2 Resultados de segunda experimentación	93
Tabla 4.3 Resultados de tercera experimentación	95
Tabla 4.4 Resultados de cuarta experimentación	96
Tabla 4.5 Experimentación de los algoritmos de machine learning con diferentes sets de datos	97

Índice de figuras

Figura 1.1: Metodología para el desarrollo de la tesis	19
Figura 2.1 Clasificación de las DEA	23
Figura 2.2 Proceso de lectura en personas con dislexia fonológica. Adaptado del modelo de lenguaje, lectura y escritura propuesto por Ellis y Young (1998)	30
Figura 2.3 Proceso de lectura en personas con dislexia superficial. Adaptado del modelo de lenguaje, lectura y escritura propuesto por Ellis y Young (1998)	31
Figura 2.4 Identificación oportuna de las DEA	42
Figura 2.5 Clasificador de machine learning para identificar un animal	45
Figura 2.6 Aprendizaje supervisado basado en etiquetas para identificar correo spam	47
Figura 2.7 Modelo de clasificador k-vecinos cercanos	48
Figura 2.8 Modelo de máquina de soporte vectorial (SVM)	49
Figura 2.9 Modelo de árboles de decisión	51
Figura 2.10 Modelo de bosques aleatorios	52
Figura 2.11 Modelo agrupación k-means	54
Figura 2.12 Modelo agrupación jerárquico	56
Figura 2.13 Arquitectura cliente-servidor	58
Figura 2.14 Ejemplificación del front end y back end	59
Figura 2.15 Informe de usuario ofrecido por la aplicación Deslixate	61
Figura 2.16 Ejercicio dentro de la aplicación Dytective	63
Figura 3.1 Arquitectura del sistema web con implementación de algoritmo de machine learning ..	66
Figura 3.2 Diagrama de contexto nivel 0 para sistema web	67
Figura 3.3 Diagrama de contexto nivel 1 para sistema web	68
Figura 3.4 Diagrama de casos de uso para sistema web	69
Figura 3.5 Diagrama de clases para sistema web	75
Figura 3.6 Diagrama de secuencia de asignación de categoría a un alumno a través del algoritmo de machine learning	76

Figura 3.7 Diagrama de secuencia de elaboración de ejercicios por parte del alumno	77
Figura 3.8 Diagrama de componentes y despliegue para el sistema web	78
Figura 3.9 Diagrama de actividades para el registro de alumnos en el sistema web	79
Figura 3.10 Diagrama de actividades para elaboración de ejercicio del alumno en el sistema web .	79
Figura 3.11 Diagrama de actividades para la asignación de una categoría a un determinado alumno registrado	80
Figura 3.12 Esquema relacional de la base de datos para el sistema web	81
Figura 3.13 Esquema general utilizado para la selección de un algoritmo de machine learning	82
Figura 4.1 Formato de alfabetización implementado por la SEC en Sonora	84
Figura 4.2 Imagen de la encuesta para recopilación de información	85
Figura 4.3 Set de datos normalizado para entrenamiento de algoritmo	88
Figura 4.4 Diferencias entre underfitting y overfitting en el proceso de entrenamiento de un algoritmo de machine learning	89
Figura 4.5 Representación de la matriz de confusión	90
Figura 4.6 Set de datos de entrenamiento	91
Figura 4.7 Visualización el árbol de decisión con el conjunto de datos final usando Orange	98
Figura 4.8 Código para exportar el modelo entrenado de árboles de decisión	99
Figura 4.9 Instrucción para activar el servidor de Laravel	101
Figura 4.10 Pantalla de inicio de sesión satisfactoria	102
Figura 4.11 Vista de los alumnos evaluados por el algoritmo de machine learning	103

Capítulo 1

Introducción

El National Joint Committee for Learning Disabilities define a las dificultades específicas de aprendizaje como un grupo heterogéneo de trastornos que se manifiestan en dificultades de adquisición y uso de las habilidades de escucha, habla, lectura, escritura y cálculo (NJCLD, 1994), que suelen manifestarse en dislexia, disgrafía y discalculia. Esta definición se ve materializada en los criterios de diagnósticos recogidos en dos de los principales sistemas de diagnósticos internacionales: la CIE-10 Clasificación Estadística Internacional de Enfermedades y Problemas de Salud (OMS, 2001) y el DSM-4-TR Manual Diagnóstico y Estadístico de los Trastornos Mentales Texto Revisado (DSM-IV-TR, 2002).

En México, entre el 6 y 8 por ciento de los niños que cursan un nivel de estudios básicos presentan algún problema o déficit de atención, afirmación realizada por Patricia Bermúdez Lozano, especialista en dislexia e investigadora de la Facultad de Psicología de la Universidad Nacional Autónoma de México. Además, se considera que entre 6% y 17% de la población infantil padece dislexia, esto afirmado por el Dr. José Benjamín Guerrero López, coordinador de la clínica del programa de salud mental de la facultad de medicina de la UNAM (Guerrero J.B., 2011).

Desafortunadamente, en nuestro país no existe un porcentaje exacto de personas que padecen algún tipo de trastorno derivado de una dificultad de aprendizaje, siendo uno de los principales factores la falta de conocimiento de los padres de familia, los altos costos de estudios para detectarlos y el hecho de que estos trastornos suelen venir acompañados de trastornos de déficit de atención y

déficit de atención e hiperactividad por sus siglas TDA y TDAH y trastornos del desarrollo de la coordinación (Artigas Pallarés, 2002); y finalmente, la manifestación de los mismos suelen estar relacionados con problemas de bajo rendimiento académico, pobre comprensión lectora, provocando en el niño una situación de desinterés escolar y bajas calificaciones. Tan solo en España, 4 de cada 10 niños con dislexia terminan abandonando sus estudios (Escribano & Enrile, 2017).

Actualmente existen una gran variedad de métodos para identificar la presencia de algunos trastornos relacionados al aprendizaje como la aplicación de pruebas de conciencia fonológica, pruebas de decodificación, pruebas de fluidez, pruebas de comprensión lectora y pruebas de nombramiento automatizado rápido, con las cuales es posible determinar el nivel de trastorno. Hoy en día, en problemas relacionados al aprendizaje de la lectoescritura es común que muchos especialistas utilicen herramientas como el DST-J, documento que contiene una serie de ejercicios que permiten evaluar de una manera empírica la dislexia (Fawcett et al., 2013) y el TOWL-4 que son pruebas de vocabulario, ortografía, puntuación implementadas para evaluar la escritura y habilidades motoras.

Por otra parte, el avance tecnológico ha permitido que se desarrollen programas o aplicaciones de apoyo a niños y personas diagnosticadas, estas herramientas suelen ser ejercicios y actividades de motivación al usuario, actividades de lectoescritura y conciencia fonológica a través de un entorno lúdico y de aprendizaje motivacional, en casos particulares, ofreciendo pre diagnóstico y posibles trastornos relacionados.

El sistema propuesto en la presente investigación consta de la implementación de un algoritmo de aprendizaje automático que incluido en un sistema web que recomiende ejercicios de lectoescritura a niños con dificultades específicas de aprendizaje, ofreciendo actividades categorizadas según los

niveles de iniciación y avances de lectoescritura del niño, apoyando su regularización durante sus primeros tres años de educación básica.

1.1 Antecedentes

Según los especialistas, los principales síntomas que caracterizan a las dificultades específicas de aprendizaje suelen derivarse de dificultades constantes en la lectura, escritura, aritmética y el razonamiento matemático, donde es posible identificar al niño con una escritura inexacta y bastante lenta, que en ocasiones suele incluso provocarle esfuerzo durante su realización. En el caso particular de problemas de lectoescritura se identifican los siguientes: letra ilegible, faltas de ortografía, comprensión lectora pobre, confusión de palabras parecidas fonéticamente, problemas de pronunciación, incapacidad para deletrear, pobre expresión escrita y letra amorfa.

Es posible encontrar aplicaciones en internet que ofrezcan ejercicios de apoyo a personas con alguna dificultad, aunque la mayoría de estos suelen ser bastantes generales ante las dificultades, no ofreciendo actividades específicas según el grado de dificultad que presenta la persona, y no contar con el apoyo de un tutor para el desarrollo. Una de las más utilizadas y que comparte bastantes características con el presente proyecto es Dytective, aplicación móvil que implementa machine learning para pre diagnosticar casos de dislexia, ofreciendo alrededor de 42000 ejercicios de lectoescritura. Dytective fue creada por un grupo de investigadores de España dirigidos por la Dra. Luz Rello, los cuales trabajan bajo una organización sin fines de lucro llamada changedyslexia, el enfoque de la aplicación es en entornos familiares, terapéuticos y pedagógicos mediante diferentes ejercicios de apoyo a niños y jóvenes. (Rello, 2018).

El segundo proyecto, llamado ModMath, es un programa adaptativo para ayudar a estudiantes con problemas de notación matemática. La aplicación permite escribir y resolver problemas matemáticos a la vez que se enfoca en trabajar con personas con disgrafía. Fue diseñada por la fundación Christopher Way y trabaja sobre la plataforma de Apple. En la biografía de los autores, se menciona que el desarrollo de la aplicación se basó en brindar apoyo a su hijo, el cual, padece una dificultad específica de aprendizaje y necesitaba de actividades más específicas para mejorar.

Finalmente, una aplicación móvil llamada Deslixate, la cual permite pre diagnosticar hasta tres tipos de dislexia, diseñada para implementarse a niños de 7 a 12 años. Este proyecto se encuentra dirigido por los mexicanos Sandra Cadena y Cesar López, que trabajando en conjunto con la UNAM y la Facultad de Estudios Superiores de Aragón (Cadena S. y López C., 2018), quienes lograron ponerlo a disposición de usuarios a través de la plataforma Android, ofreciendo la posibilidad de pre diagnosticar la dislexia a través de juegos educativos.

Actualmente, es más común encontrar instituciones educativas que ofrecen actividades de apoyo para los alumnos que presentan algún problema de aprendizaje y aunque las actividades son efectivas, en ocasiones el número de alumnos y avance de cada uno por grupo no permite que se trabajen de una manera óptima por lo cual siempre será necesario disponer de dispositivos como computadoras y tabletas que refuercen la tarea del docente frente al alumno.

Con base en lo anterior, la presente investigación partirá de una exhaustiva recopilación de información sobre los algoritmos de aprendizaje automático para recomendación, además de la identificación de aquellas actividades que pueden ser incluidas en el sistema implementado, a lo anterior, es posible sumar estrategias implementadas dentro del salón de clases que ayuden a identificar aquellos datos que serán utilizados en la implementación del algoritmo de aprendizaje

automático y que permitan una mejor predicción de las actividades a realizar para los niños dentro de sus primeros tres años de educación básica.

1.2 Definición del problema

Las escuelas de educación básica en México presentan alumnado con dificultades específicas de aprendizaje, y aunque en la mayoría se busca realizar una adecuada intervención a través de actividades que eviten lo más posible llegar a un rezago por parte del alumno, no siempre se logran los objetivos debido a la atención tan personalizada que debe tener cada uno de los niños diagnosticados debido al número de alumnos por grupo, la imposibilidad del docente para fungir como un tutor sombra sobre aquellos alumnos con necesidades mayores y la falta de material especializado para cubrir las actividades. De lo anterior, es posible determinar las siguientes preguntas de investigación:

- ¿Cuáles son las técnicas de enseñanza a niños con problemas de lectoescritura y conciencia fonológica?
- ¿Cómo utilizar la tecnología para apoyar la regularización del niño durante su proceso de alfabetización y lectoescritura?
- ¿Qué beneficio existe implementar la tecnología para apoyar al niño en sus procesos de alfabetización y lectoescritura?
- ¿Qué algoritmo de aprendizaje automático se podrá emplear para apoyar progresivamente el avance a niños con dificultades específicas de aprendizaje?

Lo anterior lleva al siguiente planteamiento del problema que guiará esta investigación:

¿Qué características se deben tomar en cuenta para la implementación de un algoritmo de aprendizaje automático que apoye en la recomendación de ejercicios a niños con dificultades específicas de aprendizaje?

1.3 Objetivos

Objetivo general: Implementar un algoritmo de aprendizaje automático en un sistema web para apoyar a niños con dificultades específicas de aprendizaje en la solución de ejercicios de lectoescritura y conciencia fonológica durante sus primeros tres años de educación básica, a través de actividades enfocadas a los procesos de alfabetización y lectoescritura.

Objetivos específicos:

- Identificar los diferentes tipos de dificultades específicas de aprendizaje.
- Analizar las actividades enfocadas a los procesos de alfabetización y lectoescritura en niños.
- Seleccionar un algoritmo de aprendizaje automático que clasifique información.
- Diseñar la arquitectura de un sistema web que implemente el algoritmo de aprendizaje automático.
- Desarrollar un sistema web que utilice un algoritmo de aprendizaje automático para la asignación de actividades enfocadas a los procesos de alfabetización en niños con dificultades específicas de aprendizaje.

1.4 Justificación

La correcta implementación de un algoritmo de aprendizaje automático que apoye en la asignación de actividades de lectoescritura a niños con problemas específicos de aprendizaje durante sus primeros tres años de educación básica puede ser una gran herramienta auxiliar tanto para el docente como para el alumno ya que será posible clasificar al alumno según sus necesidades permitiéndole elaborar actividades que consoliden su proceso de alfabetización, además de apoyar al docente que en ocasiones le suele ser complicado dedicar más tiempo a un alumno por encima del resto, ya que el número de alumnos en escuelas públicas suelen ser altos, siéndole imposible al docente dedicar el tiempo a cada uno de ellos. Por otra parte, puede ser un auxiliar en el trabajo de aula del docente, debido al apoyo como herramienta de refuerzo a niños sin un diagnóstico específico y niños ya regularizados en sus procesos de alfabetización.

De lo anterior, es necesario que las aulas de medios ofrezcan más herramientas que apoyen los procesos de regularización de cualquier alumno, sobre todo, para aquellos que presentan un determinado problema de aprendizaje.

1.5 Limitaciones y alcances

Un algoritmo de aprendizaje automático se basa en la capacidad de ofrecerle al usuario una experiencia más apegada a sus necesidades, esto, debido a que un algoritmo ha sido entrenado para ofrecer resultados basado en la experiencia aprendida del alumno a través de la información con la que ha sido entrenado y puesto en funcionamiento, por lo tanto es importante limitar nuestro alcance solamente a la implementación del algoritmo de aprendizaje automático únicamente para aquellos niños que cursan sus primeros tres años de educación primaria, dado que es en los primeros años donde el niño es alfabetizado a través de las actividades pedagógicas y refuerza sus procesos de

lectoescritura dentro del aula. Finalmente, este proyecto se desarrollará en el laboratorio de sistemas inteligentes del área de posgrado del Instituto Tecnológico de Hermosillo.

1.6 Metodología

Para el desarrollo del proyecto, se busca implementar una metodología dividida en tres fases: La primera fase hace referencia a todo el fundamento teórico del tema, la segunda fase hace referencia al análisis y diseño del algoritmo y la tercera fase que consiste en la implementación del algoritmo de aprendizaje automático a través de un sistema web y su posterior evaluación de resultados obtenidos.

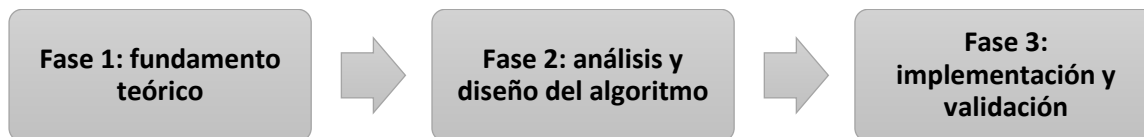


Figura 1.1: Metodología para el desarrollo de la tesis

Fundamento teórico: En esta etapa, el objetivo es recabar toda la información teórica y datos científicos sobre las diferentes dificultades específicas de aprendizaje, investigando a fondo los síntomas y tratamiento que permitan delimitar las necesidades de los niños para poder realizar una implementación adecuada del algoritmo de aprendizaje automático. Con esto, se tiene una mejor visión del objetivo general, logrando entender e identificar las principales actividades de lectoescritura y conciencia fonológica.

El realizar una investigación y un análisis profundo de las dificultades específicas de aprendizaje, así como de las técnicas actuales de enseñanza, permitirá reforzar nuestro objetivo y dar un respaldo en el desarrollo del presente proyecto.

Análisis y diseño del algoritmo: Esta fase del proyecto está conformada por el análisis y el diseño del algoritmo de aprendizaje automático, esto conlleva en primer lugar investigar los

algoritmos de aprendizaje automático más utilizados e identificando la forma en la que trabajan en la predicción de resultados y que lleve a obtener un nivel de predicción adecuado para las actividades que va a desarrollar cada uno de los alumnos a trabajar con el sistema web. Una vez que se identifiquen y estudien los algoritmos que pueden ser más adecuados, se procederá a recolectar datos que puedan permitir diseñar un algoritmo adecuado a las necesidades de los procesos de alfabetización realizando las pruebas necesarias para obtener un porcentaje alto de predicción.

Implementación y validación: Finalmente, se llevará a prueba el algoritmo de aprendizaje automático en un sistema web registrando alumno y sometidos a la predicción para que se le sea asignado un determinado grupo de ejercicios a desarrollar, donde se busca validar los resultados de la predicción lo más acertado posible a la realidad de aprendizaje del alumno.

1.7 Organización de la tesis

En el capítulo 2 se presenta el estado del arte haciendo un exhaustivo análisis de los temas más importantes de esta investigación, así como todos aquellos trabajos actuales que se encuentran relacionados con tecnologías en el apoyo a niños con dificultades específicas de aprendizaje. Se pretende revisar investigaciones que permitan determinar los ejercicios y estrategias pedagógicas actuales que se implementan en las aulas para el apoyo en el proceso de alfabetización de los niños durante sus primeros tres años de educación básica, además, se revisarán los principales algoritmos de aprendizaje automático utilizados y sus procesos de entrenamiento, validación e implementación.

El capítulo 3 describe el análisis y el diseño del algoritmo de aprendizaje automático a implementar en un sistema web. Se presentan los diseños y esquemas a trabajar en el desarrollo del

algoritmo de aprendizaje automático, así como la propuesta de arquitectura del sistema web donde se llevará a cabo su implementación.

El capítulo 4 describe los procesos de experimentación del algoritmo de aprendizaje automático y el desarrollo de su sistema web, además de presentar las diferentes pruebas realizadas para la correcta selección del algoritmo a utilizar en un entorno de producción.

En el capítulo 5 se describen y analizan los resultados obtenidos de la implementación del algoritmo, así como las diferentes pruebas del sistema para la correcta compatibilidad en un entorno web.

Finalmente, en el capítulo 6 se presentan las conclusiones obtenidas durante el desarrollo del trabajo y se plantean los posibles trabajos a futuro generados a partir de esta investigación.

Capítulo 2

Estado del arte

El presente capítulo inicia abordando los conceptos y trastornos que forman parte de las dificultades específicas del aprendizaje, explicando brevemente su manifestación y sus procesos de diagnósticos, abordando los principales detalles de cada una de ellas, así como el impacto en la educación.

Se abordan los conceptos más relevantes de la inteligencia artificial, enfocándola hacia un área conocida como el aprendizaje automático o machine learning, explicando los principales algoritmos utilizados para la predicción de información.

Finalmente, se anexan conceptos básicos sobre el algoritmo a utilizar y cómo llevar a cabo la implementación en un entorno de trabajo web, esto, debido a la necesidad de tener acceso al sistema desde cualquier dispositivo.

2.1 Dificultades específicas del aprendizaje

Las dificultades específicas del aprendizaje es un término que hace referencia a un grupo heterogéneo de alteraciones que se manifiestan en dificultades en la adquisición y uso de habilidades de escucha, habla, lectura, escritura y razonamiento matemático. Dichas alteraciones son intrínsecas en el individuo, es decir, permanecerán con él durante su periodo de vida. Estos trastornos son clasificados en el DSM-IV como problemas con el reconocimiento de palabras de forma precisa o fluida, mal deletreo y poca capacidad ortográfica (DSM-IV-TR, 2002).

Las DEA como dificultades en las áreas primarias de aprendizaje (lectura, escritura y cálculo), se encuentran agrupadas en tres trastornos: la dislexia, la disgrafia y la discalculia respectivamente, las cuales se ilustran en la figura 2.1 y se definen en los siguientes párrafos.

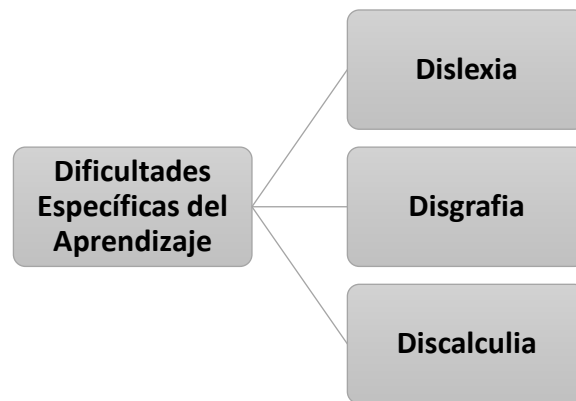


Figura 2.1: Clasificación de las DEA

2.1.1 Dislexia

La definición de dislexia ha sido objeto de debate a lo largo de la historia originando el surgimiento de múltiples definiciones de la misma. Los primeros informes de carácter científico sobre la dislexia se remontan a 100 años atrás en el ramo de la medicina y la oftalmología, siendo un científico alemán el primero en utilizar el término “dislexia” para describir las dificultades de pacientes adultos que habían sufrido un daño cerebral (Hoiem, n.d.). En el año de 1896, el pediatra P. Morgan fue el primero en describir un caso de dislexia infantil, con un chico llamado Percy, que en ocasiones escribía su nombre “Precy”, utilizando el término “ceguera de palabras” (word blindness, en inglés) para hacer referencia a la dislexia, y así, a través de los años se dio comienzo al estudio en niños con problemas de lectura, escritura y lenguaje (Morgan, 1896).

Fue en la década de los 60, cuando Donald Critchley acuñó el término dislexia del desarrollo, defendiendo la opinión que representaba un síndrome neurológico. En ese entonces, comienza a

delimitarse la dislexia como un trastorno específico del aprendizaje de la lectoescritura, es decir, la presencia en el individuo de una dificultad que no se explica por las capacidades cognitivas generales, ni por otras razones, como una instrucción inadecuada, problemas sensoriales o problemas emocionales severos (Critchley, 2015).

Es en tiempos más recientes donde se adopta la definición de dislexia por parte de la Asociación Internacional de Dislexia, describiéndola como un trastorno específico del aprendizaje cuyo origen neurobiológico es caracterizado por dificultades en el reconocimiento preciso de las palabras (escritas) y déficit en la decodificación (lectora) y deletreo. (IDA, 2002).

De acuerdo con la definición anterior, la persona que padezca dislexia presentaría, además, reducción en el vocabulario y una escasez del conocimiento verbal, dificultades que serían secundarias al déficit de comprensión lectora y reducción en experiencia lectora. Generalmente, entre los alumnos que presentan problemas de lectura se encuentran dos grupos diferenciados: alumnos que leen correctamente las palabras, pero no comprenden el texto y alumnos que tienen problemas para leer las palabras, como consecuencia, mostrando dificultad para extraer el significado global del texto. A su vez dentro de este último grupo, se pueden distinguir entre disléxicos y lectores atrasados (Jiménez, 2012).

En el caso de la dislexia, se tiene un trastorno específico de la lectura que se manifiesta en personas que teniendo intacta su comprensión oral, presentan sin embargo problemas en el reconocimiento de palabras escritas; de igual modo debe entenderse que dichas dificultades no son consecuencia de problemas emocionales, problemas sensoriales, problemas socioculturales, escasa escolarización o incluso bajo coeficiente intelectual, pues es sabido que dicho trastorno no afecta la inteligencia de la persona, caso contrario, los niños con retraso lector y bajo coeficiente intelectual,

leen mal, pero además, presentan dificultades generalizadas en el aprendizaje, en cambio los niños con dislexia dan muestra de una capacidad intelectual normal o superior, aunque compartan en este grupo su bajo nivel lector.

De lo anterior, podemos decir que la dislexia es un trastorno y no una enfermedad, es decir, se presenta una alteración afectando la funcionalidad de la conducta lectora impidiendo poder extraer de manera correcta la información escrita, afectando al individuo en su adaptación académica, personal y social.

La persona que presenta dislexia, podrá tener al menos los siguientes síntomas durante su proceso de aprendizaje: comprensión lectora baja, errores en la lectura y escritura como son omisión, sustitución y trasposición de números, letras y palabras; errores gramaticales en la escritura y ortografía deficitaria, dificultad en conectar grafemas (letras) con fonemas (sonidos) lo que consecuentemente ocasionará inseguridad en la pronunciación de las letras, dificultad para rimar, dificultad para acceder a su léxico como es nombrar números o letras, problemas conductuales por falta de atención e impulsividad, problemas de autoestima o dificultad para integrarse, problemas con las matemáticas (algunos casos) y un desarrollo asimétrico, es decir, puede presentar gran habilidad manual o gran creatividad en comparación con su habilidad lingüística (Rello, 2018).

A pesar de los síntomas que puede llegar a padecer, es importante aclarar que la dislexia no es un síntoma de otro, sino un trastorno en sí mismo, sin embargo, puede aparecer junto con otros trastornos del aprendizaje, del lenguaje y del comportamiento, ya que existe una gran comorbilidad entre trastornos del desarrollo como lo son trastornos específicos del lenguaje (TEL) o retraso simple del lenguaje (RSL), también puede acompañar a trastornos como el TDA/TDAH. Este último, a pesar

de no ser una dificultad específica del aprendizaje afecta una habilidad cognitiva general, que es la atención o el control de la conducta del individuo (Snow et al., 2015).

2.1.1.1 Clasificación de la dislexia

A través de los años se han venido proponiendo diversas clasificaciones de la dislexia, la primera distinción que cabe hacer, es entre dislexia adquirida y dislexia del desarrollo (evolutiva), la primera es consecuencia de lesiones cerebrales adquiridas en una determinada zona de la corteza a la que se le atribuye la función quedando alterada, mientras que la última suelen ser más comunes y sobre las cuales se han venido trabajando a través de los años.

Los primeros trabajos sobre la variabilidad de la dislexia provienen de aquellos sobre los cuales se buscó encontrar o clasificar subtipos de dislexia. (Snowling, 2000) fue de los pioneros en realizar revisiones de distintas clasificaciones de dislexia proveniente de varios autores que ya venían trabajando con anterioridad este tema. Además, ante el surgimiento de teorías sobre la dislexia surgieron múltiples clasificaciones, sin embargo, cabe distinguir entre la dislexia adquirida y la evolutiva (Farnham-Diggory, 2004):

- Dislexia adquirida: causada a raíz de una lesión cerebral afectando las áreas que participan en el proceso de la lectoescritura
- Dislexia evolutiva: aquella que muestran los niños con dificultad en el aprendizaje de la lectura

Otra clasificación que es aceptada es aquella basada en las rutas con acceso al componente léxico afectado (Farnham-Diggory, 2004):

- Dislexia fonológica: producida por un problema en la ruta fonológica
- Dislexia visual: producida por un problema en la ruta visual

- Dislexia mixta: tanto la ruta fonológica y la ruta visual se encuentran afectadas dificultando el procesamiento de la información.

Aunque las clasificaciones de la dislexia han variado a través de los años, es importante mencionar que ninguna de ellas se ha impuesto, y esto es debido a la dificultad de demostrar cuáles son las alteraciones básicas que condicionan a la dislexia, además de venir acompañada de otros trastornos, lo que en numerosas ocasiones hace que no sea detectada en forma. La tabla 2.1 resume diversas clasificaciones de la dislexia a través del tiempo, donde es posible distinguir diferentes definiciones para referirse al mismo concepto o uno similar.

Tabla 2.1: Clasificaciones de la dislexia propuestas por Josep Artigas

Autores	Clasificación
Border (1973)	Dislexia disfonética Dislexia diseidética Dislexia mixta
Bakker (1979)	Dislexia L (Dislexia lingüística) Dislexia P (Dislexia perceptiva) Dislexia M (Dislexia mixta)
Mattis (1975)	Dislexia con alteración primaria del lenguaje Dislexia con trastorno articulatorio-grafomotor Dislexia con trastorno visuoperceptivo
Baddeley (1982) Coltheart (1983) Temple (1983) Marshall (1984)	Dislexia superficial Dislexia fonológica Dislexia profunda

Se dice que solo hay un tipo oficial de dislexia, sin embargo, se puede encontrar literatura de carácter científico que habla de diferentes tipos que hasta el día de hoy siguen siendo debatidos, pero es importante remarcar que en las diferentes definiciones científicas de dislexia se encuentran dos cosas en común: 1) se trata de un problema del lenguaje escrito, no del lenguaje oral; 2) no se encuentra relacionado con otras habilidades cognitivas, dicho de otra forma, no está relacionado con la inteligencia.

2.1.1.1.1 Dislexia del desarrollo

También conocida como dislexia evolutiva o dislexia específica de evolución (según los autores), su principal característica es la predisposición a desarrollarla desde su nacimiento, ya que, en la mayoría de los estudios, se sabe que trastornos de este tipo, suelen ser hereditarios. La dislexia se nota cuando por primera vez, el niño ingresa a preescolar, a su primer año de educación primaria o cuando se encuentra frente al desafío de la lectoescritura.

Abordando el tema de la lectura, para leer es necesario captar la correspondencia que existen en los símbolos o palabras que utilizamos para representarlos (grafemas) y los sonidos del lenguaje (fonemas). Es importante dominar esta habilidad, especialmente durante el periodo de aprendizaje de la lectura, pero también más tarde, cuando el lector adulto debe leer palabras irregulares y pseudopalabras, es decir, aprender a identificar palabras irregulares que se pronuncian distinto a como se escriben como por ejemplo “house” y las pseudopalabras, aquellas palabras que son conjunto de letras pronunciables, pero sin ningún significado en el idioma, como por ejemplo “casinota”; esto, para poder separar los distintos procesos fonológicos, auditivos y visuales involucrados en la lectura (Galaburda & Cestnick, 2003).

De lo anterior, la lectura de pseudopalabras requiere de procesos fonológicos y auditivos, pero no de procesos léxicos, siendo este último, el que aporta más ayuda en el proceso de lectura de palabras irregulares. Esto es debido a que las pseudopalabras se leen tras aplicar los procesos de pronunciación, mientras que para las irregulares no ocurre lo mismo, debido a la forma en la que se pronuncia con respecto a su escritura.

Con lo anterior, es posible clasificar los dos principales subtipos de dislexia de desarrollo: la dislexia fonológica, como un problema para leer pseudopalabras, un trastorno de la lectura en método no léxico; y la dislexia superficial, como un problema para leer palabras irregulares, un trastorno de la lectura de palabras complejas.

2.1.1.1.1 Dislexia fonológica

En personas con dislexia fonológica, la conversión de grafema-fonema, da como resultado una lectura con numerosos errores en la decodificación además de baja velocidad en su proceso. Suelen presentarse demasiadas correcciones durante la misma, sustituyendo algunos grafemas por otros con semejante fonología y aspecto visual (p-d, p-q, q-b, b-d, m-n, t-d, b-t, p-t, k-g, m-b), además de confusiones entre grafemas dependientes del texto (c, g) y grafemas de baja frecuencia (x, j, k). También se pueden observar los efectos de la frecuencia léxica donde la lectura mejora con las palabras familiares, pero existiendo mayor dificultad a medida que desciende la familiaridad de palabras, con un número mayor de errores en palabras cortas que largas.

También se observan los efectos de la frecuencia léxica y su familiaridad, de tal modo que la lectura suele mejorar ante palabras familiares, presentándose mayor dificultad en la medida que disminuye la familiaridad de las palabras y aumento en la dificultad ante nuevas palabras (Davies et al., 2007).

Las personas con dislexia fonológica solo son capaces de leer mediante su ruta léxica, ya que se encuentra alterada su ruta fonológica, debido a la incapacidad de implementar los métodos de conversión de grafema-fonema caracterizando por leer bien las palabras más familiares ante la ausencia de pseudopalabras o palabras desconocidas; comúnmente llegan a cometer errores visuales

en pseudopalabras que se parecen a palabras con abundantes lexicalizaciones, ejemplo de ello: antiguo por artiguo, playa por blaya, espada por espaca y durante la lectura presentan errores en palabras parecidas, ejemplo de ello: firme por forma. Debido a esto, las personas con dislexia fonológica presentan principalmente este tipo de problemática, como una consecuencia de sus problemas de recodificación fonológica, ya que el procesamiento fonológico de las palabras facilita la formación del léxico mental (Share & Stanovich, 1995), podemos observar la figura 2.2, la cual muestra un mejor detalle del proceso de lectura para aquellas personas diagnosticadas con dislexia fonológica.

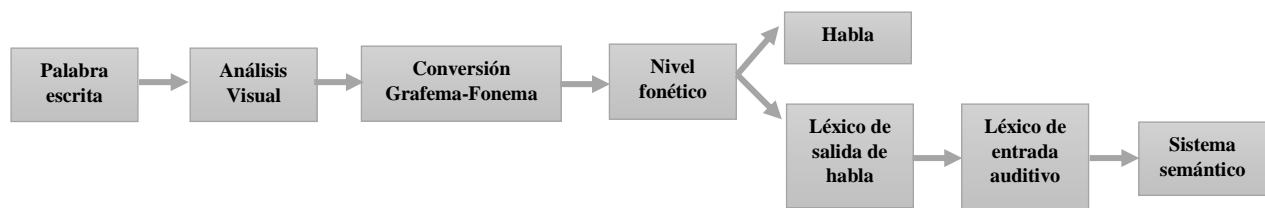


Figura 2.2 Proceso de lectura en personas con dislexia fonológica. Adaptado del modelo de lenguaje, lectura y escritura propuesto por Ellis y Young (1998)

2.2.1.1.1.2 Dislexia superficial

Las personas con dislexia superficial pueden leer a través de su proceso fonológico, pero no del léxico, debido a ello, normalmente suelen ser incapaces de reconocer una palabra de forma léxica. El bajo reconocimiento de palabras produce una lectura bastante defectuosa, fragmentada y lenta, esto, consecuente al lento delecteo, la mala prosodia y el bajo ritmo de lectura, aunque en la mayoría de los casos, con pocos errores durante la misma.

La dislexia superficial se presenta en la lectura de palabras irregulares, es decir, palabras que se escriben de una manera y se pronuncian de otra; aunque lo anterior no presenta gran problema en idiomas como el español y el italiano (M Coltheart et al., 2001), ya que son considerados bastantes regulares y sus palabras son leídas con base en su sistema fonológico-auditivo y no por el léxico,

mientras que en idiomas como el inglés, la mayoría de las palabras son leídas con base en su sistema léxico-visual. A raíz de esto, es normal encontrar más casos de dislexia superficial en el idioma inglés que en el idioma español (Zorzi et al., 1998); pudiendo hacer mención para aclarar lo anterior, el ejemplo de un niño pre letrado, cuando llega a ver algún logotipo comercial o alguna marca en específico, logra identificarla debido a su sistema léxico, pero inhibiendo su sistema fonológico, ya que este último no ha madurado lo suficiente.

Es importante dejar en claro, que todas las personas disléxicas presentan dificultados de lectura en ambos tipos de palabras (pseudopalabras e irregulares), pero el desempeño en cada tipo de dislexia es mucho más inferior al del otro, aunque la mayoría de los afectados presentan mayor dificultad en la lectura de palabras irregulares, lo que significa la presencia de dislexia fonológica mayormente. Lo anterior, se detalla a través de la figura 2.3, que muestra el proceso de lectura en personas con dislexia superficial.

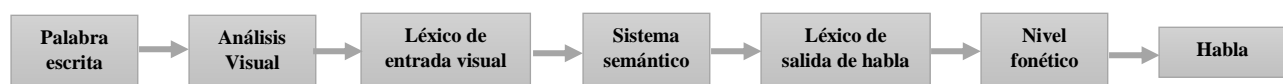


Figura 2.3: Proceso de lectura en personas con dislexia superficial. Adaptado del modelo de lenguaje, lectura y escritura propuesto por Ellis y Young (1998)

2.1.1.1.2 Dislexia adquirida

La dislexia es un problema permanente y persistente en el lenguaje escrito de las personas, si bien, este problema puede ser hereditario, existen situaciones donde personas que ya saben leer y

escribir pueden presentarla, debido principalmente a algún daño cerebral en las áreas donde se da el procesamiento de lectoescritura o cuando logra manifestarse en personas, en este caso, niños que aún no han aprendido a leer ni escribir, causándole grandes dificultades en su proceso de aprendizaje, cuyo origen pudiera estar relacionado con deficiencias constitucionales, siendo ambas situaciones una dislexia adquirida.

Una de las clasificaciones más clásicas de los disléxicos es la propuesta por Castles y Coltheart (1993), quienes desarrollaron su tipología basándose en investigaciones en problemas de lenguaje consecuentes a un daño cerebral, teniendo como referencia el marco explicativo del modelo de doble ruta.

La idea del modelo de doble ruta (léxico y no léxico), deriva de la llamada “teoría de la vía dual” de la lectura, esta misma, tiene sus orígenes en el análisis de casos de lesiones cerebrales adquiridas posteriores al aprendizaje de la lectura, que demuestran trastornos específicos de una u otra vía y que depende de la localización de la lesión en el mismo cerebro (Max Coltheart & Leahy, 1996).

Este tipo de dislexia suele ser el resultado de un traumatismo o lesión en la parte del cerebro que controla la lectura y escritura o bien, puede ser resultado de un tumor cerebral o un accidente cerebrovascular. Los pacientes con dislexia adquirida pierden la capacidad para leer con fluidez, como se mencionó anteriormente debido a una lesión en la parte posterior del hemisferio izquierdo del cerebro, curiosamente, los pacientes con este tipo de dislexia, suelen caminar, hablar, moverse con facilidad e incluso pensar con normalidad previo a su lesión, pero no saben leer ni escribir, incluso en algunos casos ni lo que ellos mismos escriben; de modo que en situaciones donde existieron casos de infarto cerebral, y en otros ejemplos de casos de lesiones adquiridas, el lugar del daño que produce

trastornos de la lectura de pseudopalabras es distinto al sitio del daño que produce problemas con la lectura de palabras irregulares (Share & Stanovich, 1995).

En resumen, los casos de dislexia fonológica se producen tras daños en el giro temporal superior y regiones temporoparietales del cerebro, mientras que, en la dislexia superficial, los daños se asocian más a regiones visuales de asociación y en el borde parietooccipital. Por lo tanto, debería darse el caso que en el grupo de dislexia del desarrollo se descubrieran dos mecanismos anatómicos distintos que explicasen la diferencia entre la dislexia fonología y la dislexia superficial (Small et al., 1998).

2.1.2 Disgrafía

La American Psychiatric Association describe a la disgrafia como una discapacidad en la expresión escrita y unas habilidades de escritura que están, sustancialmente, por debajo de lo que cabría esperar en relación a la edad, la inteligencia y la etapa educativa en la que se encuentra la persona (APA, 2014).

La disgrafia es la dificultad del acto de escribir y que conlleva a problemas en la ortografía, mala caligrafía y dificultades para plasmar el pensamiento en escrito. Es un trastorno específico del aprendizaje que dificulta la expresión escrita y que debe ser evaluada y diagnosticada por un psicopedagogo a partir de los 7 años, ya que antes de esa edad los niños solo presentan algunos síntomas (Ronquillo, 2016).

Según (Vega & Arévalo, 2011): “La disgrafia escolar es una dificultad que experimenta el estudiante para recordar cómo se forman determinadas letras, dificultándole el trazado de las mismas”.

Las causas que originan a la disgrafía son diversas, según Thorne (2004, citado en Farham-Diggory, 2004), puede ser debida a alteraciones en algunos de los componentes de la escritura como:

- Habilidad de interpretar lo que se ve (destreza visio-perceptiva).
- Habilidad de representar una palabra escrita en la memoria y su capacidad para recuperarla (codificación ortográfica).
- Habilidad motriz.
- Coordinación visio-motriz.

2.1.2.1 Clasificación de la disgrafía

Al igual que la dislexia, en este ámbito, podemos encontrar un número significativo de clasificaciones y nomenclaturas basadas principalmente en aspectos iconográficos y síndromes gráficos; pero es posible aportar una clasificación sencilla según Deuel (1995); (Richards, 1998), donde distinguen tres tipos de disgrafía: disgrafía disléxica, disgrafía motriz y disgrafía espacial.

2.1.2.1.1 Disgrafía disléxica

También conocida como disgrafía de procesamiento, la disgrafía disléxica se considera una proyección de la dislexia en la escritura, lo que conlleva a una escritura ilegible y una ortografía bastante deficiente, además, se puede encontrar la presencia de copiado de textos buenos, destreza para dibujar y presencia de motricidad manual de una manera normal. Según Portellano (Portellano, 2001), la persona con disgrafía disléxica presentará errores similares a la dislexia, de ahí su referencia, pero dichos errores, corresponderán a la escritura, pudiendo estar presente al menos:

- Omisión de letras, sílabas o palabras.
- Confusión de letras con sonidos semejantes.

- Confusión de letras con orientación simétrica similar.
- Inversión del orden de las sílabas, por ejemplo: sol por los, el por le.
- Separaciones indebidas de sílabas, letras o palabras.
- Exceso de rasgos en las letras, escritura en espejo y reiteración de letras.

2.1.2.1.2 Disgrafía motriz

Se habla de disgrafía motriz cuando se producen alteraciones en las capacidades motrices de la persona, provocando una afectación en la fluidez al escribir llevando a una escritura bastante deficiente, en lo que respecta a la copia de textos, se puede apreciar poca legibilidad al igual de la deficiencia en el dibujo y una motricidad manual bastante deficiente. Las principales manifestaciones en este tipo de disgrafía (Portellano, 2001) se pueden observar en:

- Trastornos en la forma de las letras, donde normalmente no se llegan a cerrar bien letras como la o, a, b, y otras similares.
- Se presentan alteraciones en los tamaños de las letras, escribiendo algunas en tamaño regular, pero a la vez que se va perdiendo la simetría de tamaño en las mismas.
- Hay un gran esparcimiento entre las letras de una palabra, pudiéndose a la vez presentar una inclinación defectuosa al escribir un renglón.
- Trastornos de fluidez en el ritmo de la escritura, provocando que en ocasiones, tarde más tiempo en escribir ciertas palabras, sin mantener un ritmo fluido.
- Intercalación de letras mayúsculas con minúsculas y posición inadecuada al tomar el lápiz para escribir.

2.1.2.1.3 Disgrafía espacial

La disgrafia espacial conlleva un problema de procesamiento visual y comprensión del espacio, es decir, este tipo de disgrafia provoca que la persona tenga una escritura espontanea con palabras desordenadas, letras separadas, escritura inclinada y problemas en el coloreado.

Las principales características de la disgrafia espacial suelen presentar los siguientes síntomas con respecto a su escritura (Hudson et al., 2017):

- No escriben recto y no son capaces de seguir una pauta.
- Presentan problemas en la organización de palabras de izquierda a derecha.
- Hay espacios incongruentes entre las letras y las palabras, es decir, puede que se encuentren demasiado cerca o demasiado lejos entre sí.
- Las letras y palabras presentan tamaño, forma e inclinación irregular.
- Presentan problemas para desplegar respuestas de tipo aritmético o para la geometría, tal como se puede observar en la tabla 2.2.

Tabla 2.2: Características diferenciales entre los distintos tipos de disgrafia.

Tipo de disgrafia	Escritura	Ortografía	Copiado de texto	Dibujo	Motricidad
Disléxica	X	X			
Motriz	X		X	X	X
Espacial	X		X	X	

Como se puede observar en la tabla anterior, la dificultad más general en las personas con disgrafía se produce en la escritura, diferenciándose de la dislexia, mientras que la única diferencia entre la disgrafía motriz y la disgrafía espacial, es la afectación de la capacidad motriz del niño.

2.1.3 Discalculia

Se puede definir como un problema de aprendizaje relacionado con las matemáticas, según Breznitz (2009), La discalculia es un trastorno específico del aprendizaje, de base biológica, que afecta profundamente el aprendizaje de las capacidades aritméticas y las matemáticas, en ocasiones se define como “dislexia matemática”. Es independiente al nivel de inteligencia del niño y también a los métodos pedagógicos empleados. La dificultad se centra en la capacidad de interpretación de los símbolos numéricos y los conocimientos aritméticos como la suma, resta, multiplicación y división.

Kosc, (1970) realizó estudios en casos escolares de niños con discalculia con el objetivo de buscar causas, síntomas y características aportando la siguiente definición:

Un trastorno estructural de habilidades matemáticas que tiene su origen en un trastorno genético o congénito que aquellas partes del cerebro que son el sustrato anatómico-fisiológico directo de la maduración de habilidades matemáticas adecuadas a la edad, sin un trastorno simultáneo de las funciones mentales generales. (p. 56).

La discalculia se manifiesta por un debilitamiento o pérdida de la capacidad de calcular, manipular símbolos numéricos o hacer operaciones aritméticas simples, normalmente a la discalculia, se le atribuye déficits verbales, espaciales, secuenciales y cognitivos, es decir, el niño puede lograr leer y escribir, pero no calcular.

De acuerdo a los criterios que se encuentra en el DSM-IV la discalculia se puede presentar si:

- La capacidad de cálculo, evaluada mediante las pruebas correspondientes, se encuentra por debajo de lo esperado según la edad de la persona, su coeficiente intelectual y la escolaridad de acuerdo a su edad.
- El trastorno en el punto anterior interfiere significativamente en el rendimiento académico o en las actividades de la vida diaria que requieran capacidades de cálculo.
- Si hay un déficit sensorial, las actividades para el cálculo exceden de las que normalmente se encuentran asociadas a él.

Además de las causas de discalculia por lesiones cerebrales, alteraciones neurológicas, se dice que también está relacionada a problemas de lateralidad, de sentido espacial, de memoria, de atención, de dificultad en la adquisición del pensamiento abstracto y la psicomotricidad (Farnham-Diggory, 2004).

2.1.3.1 Clasificación de la discalculia

La discalculia se puede clasificar en cuatro tipos según su área o aspecto matemático donde se presenta dicho trastorno (Farnham-Diggory, 2004), siendo una de las más utilizadas clasificaciones, aunque también se encuentran las propuestas por (Kosc, 1970), (Giordano, 1978). La detección actual dependerá de los especialistas, pero los tratamientos siempre se enfocarán en contrarrestar y mejorar las deficiencias aritméticas del paciente. A continuación, definiremos brevemente las clasificaciones propuestas por Kosch (1970).

2.1.3.1.1 Discalculia verbal

La discalculia verbal está representada por el problema para nombrar y comprender conceptos matemáticos a través del lenguaje (verbal), se dice que los niños son capaces de leer y escribir los números, pero no recordarlos cuando llegan a ser pronunciados por otros.

2.1.3.1.2 Discalculia léxica

Se presenta discalculia léxica, cuando la persona presenta problemas en la lectura de símbolos matemáticos, es decir, presenta gran dificultad en la lectura de símbolos, números o expresiones matemáticas, la persona es capaz de entender los conceptos matemáticos cuando alguien más habla de ellos, pero las dificultades se presentan cuando se trata de leerlos y comprenderlos.

2.1.3.1.3 Discalculia gráfica

La discalculia gráfica se da dentro de la grafía de los números, esto quiere decir el método para poder escribir o de dibujar de una manera correcta los números, símbolos y ecuaciones. Las personas que padecen este tipo de trastorno, son capaces de entender los conceptos matemáticos, pero no tienen la capacidad para leerlos y escribirlos.

2.1.3.1.4 Discalculia practognóstica

La discalculia practognóstica se presenta cuando la persona tiene dificultades en la comparación de cantidades de forma manipulativa, es decir, tiene su sustento en el proceso de traducir el conocimiento sobre conceptos abstracto-matemáticos a conceptos reales. Principalmente presentan problemas para enumerar, comparar y manipular operaciones matemáticas en la práctica.

2.1.3.1.5 Discalculia ideognóstica

Se define este tipo como la dificultad asociada a la comprensión de conceptos y sus relaciones, básicamente, las personas con este trastorno presentan dificultades para realizar cálculos u operaciones de manera mental, sin usar los números para llegar a los resultados, además de la comprensión de conceptos relacionados con la aritmética. Estas personas, presentan problemas para recordar conceptos matemáticos una vez que ya lo aprendieron.

2.1.4 Diagnóstico y tratamiento

Las dificultades específicas del aprendizaje (DEA) se manifiestan de forma distinta entre cada uno de los individuos que la padecen; no existen dos personas que tengan los mismos síntomas. En todos los casos no se consideran una enfermedad, más bien, son trastornos específicos de carácter congénito y hereditario en las personas que lo padecen.

Los principales síntomas para identificar casos de DEA (Cuetos, 2006) van a depender de factores según el rango de edad de la persona, pudiéndose presentar algunos de los siguientes:

- Retraso en el aprendizaje del lenguaje.
- Confusión en pronunciación de palabras con fonética similar.
- Falta de habilidad para recordar el nombre de serie de cosas o secuencias.
- Con función con el vocabulario que tiene mayor orientación espacial.
- Dificultad para aprender las rimas típicas en edad de pre escolar.
- Persistente tendencia a escribir números y palabras en espejo o con orientación inadecuada.
- Dificultad para distinguir izquierda y derecha.
- Dificultad para aprender el alfabeto y las tablas de multiplicar.

- Falta de atención y concentración.
- Baja autoestima, frustración, impulsividad, periodos atencionales más cortos.
- Escritura marcada por omisiones, adiciones de letras o alteraciones del orden de las mismas.
- Desorganización en casa y escuela.
- Dificultad para seguir instrucciones orales.
- Diferencia entre la estructuración y organización del pensamiento oral y escrito.
- Inconsistencias gramáticas y errores ortográficos.
- Dificultad en el aprendizaje de lenguas extranjeras.
- Aversión a la lectura y escritura.
- Torpeza y coordinación manual baja.
- Postura inadecuada, tanto del niño como de la hoja del papel.
- Tonicidad muscular inadecuada.
- No se usan los signos de puntuación para las pausas previstas.
- Lectura mecánica debido al esfuerzo en la decodificación.

La identificación temprana es importante para disponer del tiempo y tratamiento necesario en el proceso de regularización del niño, en etapas mayores a las mencionadas, de acuerdo a la figura 2.4, es posible lograr un buen diagnóstico mediante una entrevista familiar que permita identificar antecedentes y una valoración neuropsicológica, tal como se ha explicado en las anteriores dificultades, en caso de no dar la atención adecuada, los problemas se van agravando e incluso la persona puede llegar a presentar inseguridad personal y baja autoestima.

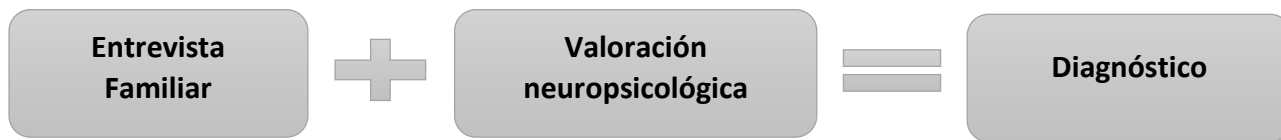


Figura 2.4. Identificación oportuna de las DEA.

De la imagen anterior, es importante considerar estudiar la presencia de antecedentes directos de primera generación de dificultades de aprendizaje en el periodo escolar, es decir, determinar el desarrollo que presentaron los padres durante sus etapas escolares y su desarrollo, pues como ya se mencionó anteriormente, las DEA y en el caso particular de la dislexia, son hereditarias y en algunos casos, nunca se logró detectar en los padres.

La valoración neuropsicológica siempre se llevará a cabo por un especialista en la materia, un logopeda o maestros especializados en trastornos del aprendizaje, neuropsicólogos y psicopedagogos, quienes, a través de diferentes pruebas, podrán evaluar la situación del niño. Los tratamientos van a depender de la edad en que sea diagnosticada la persona, principalmente se dice que entre los 6 y 9 años los objetivos siempre deberán ser enfocados a aumentar la conciencia fonológica; tanto oral como escrita, buscando mejorar la automatización de la mecánica lectora. Nótese que, en esta etapa, es cuando el niño puede presentar mayor rechazo, ya que las actividades escolares posiblemente le lleven mayor tiempo de realización y lleguen a ser agotadoras y poco gratas.

A partir de los 10 años está comprobado que es difícil el aumentar la conciencia fonológica y la automatización de la lectura, por lo tanto, los objetivos a trabajar serán diferentes, buscando la enseñanza de estrategias de comprensión de textos (búsqueda de palabras claves, subrayado, resumen, etc.).

2.2 Inteligencia artificial

La real academia de la lengua española define la inteligencia como potencial intelectual, la cual es la facultad de conocer, entender o comprender. Con esta definición, la inteligencia artificial se puede definir de la misma forma, pero con enfoque hacia las máquinas. Según Marvin Minsky, uno de los pioneros del tema, la inteligencia artificial es la ciencia de construir máquinas para que hagan cosas que si las hicieran los humanos, requerirían inteligencia.

Según la Encyclopedia of Artificial Intelligence la IA es un campo de la ciencia y la ingeniería que se ocupa de la comprensión, desde el punto de vista informático, de lo que domina comúnmente comportamiento inteligente. También se ocupa de la creación de artefactos que exhiben este comportamiento. Otros autores definen a la inteligencia artificial como el estudio de las ideas que permiten a las computadoras ser inteligentes (Winston, 1992).

Podemos pensar que la inteligencia artificial es aquella ciencia que incorpora un nivel de conocimiento a los procesos o actividades para que estas tengan un determinado éxito, dicho con un ejemplo, el ajedrez es un juego donde en cada jugada se evalúan las posibilidades tanto ofensivas como defensivas, teniendo un sinnúmero de jugadas para realizar, pues es imposible pensar que una computadora pueda evaluar todas las jugadas posibles, ya que en vez de esto, resulta mejor aplicar el conocimiento en buscar la mejor jugada a través de evaluaciones inteligentes.

En 1950, Alan Turing propuso un juego llamado “juego de la imitación”, el cual se jugaba con tres personas que no se conocen entre sí. Dos de ellas juegan como testigos y son del sexo opuesto, el tercer jugador, es un interrogador, cuya función es tratar de adivinar de cual testigo se trata, basándose solamente en cómo contestan sus preguntas. El truco del juego consiste en que uno de los testigos

trata de engañar al interrogador haciéndose pasar por la otra persona, mientras que el otro testigo hace lo posible por ayudar al interrogador. Si el interrogador adivina, gana el testigo que lo ayudo, de lo contrario, gana el testigo que lo engaña.

La idea de Turing era sustituir al testigo masculino por una computadora y ver si frente a un promedio de adversarios mujeres, podía engañar al promedio de interrogadores (hombres) con la misma frecuencia con que la podía hacer al promedio de hombres, en caso de hacer, la computadora pasaría la prueba (Haugeland, 1988).

2.2.1 Machine learning

El machine learning o aprendizaje automático, permite la implementación de algoritmos para el reconocimiento de patrones permitiendo a las computadoras la capacidad de aprender a partir de dichos modelos implementados en un set de datos; Tom Mitchell, profesor del departamento de machine learning de la universidad de Carnegie Mellon, indica que el aprendizaje automático busca responder preguntas en como estructurar sistemas informáticos que mejoren automáticamente con la experiencia.

El machine learning es una disciplina científica proveniente de la inteligencia artificial, que estudia cómo los sistemas pueden programarse para aprender y mejorar con la experiencia sin la intervención humana (Ibañez, 2018).

Los algoritmos de machine learning aprenden a partir de los datos sobre los cuales son implementados, de esta manera, las maquinas son entrenadas para que aprendan a ejecutar determinadas actividades de manera automática, así, cuando son expuestas a nuevos datos, los algoritmos se adaptarán a partir de los cálculos anteriores y los patrones se moldean para ofrecer

respuestas más confiables. En la figura 2.5 se muestra un ejemplo, donde un modelo de machine learning es capaz de identificar un animal a partir del sonido que este emite, al ser entrenado el algoritmo con una determinada cantidad de sonidos le han permitido clasificar los nuevos sonidos.

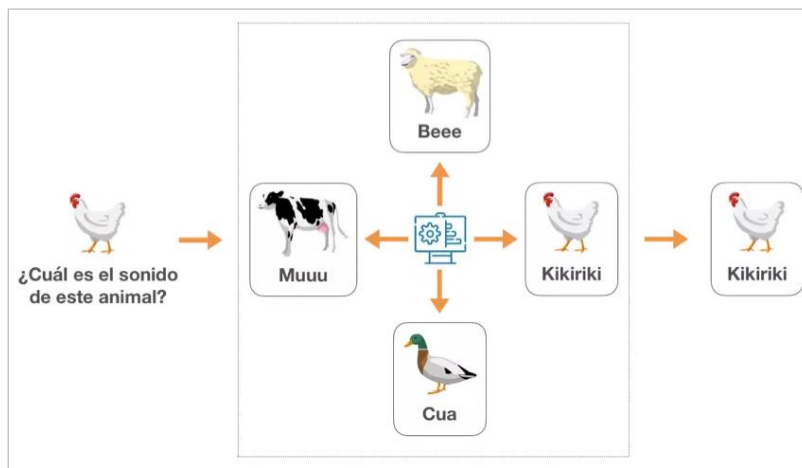


Figura 2.5: Clasificador de machine learning para identificar un animal

El principal aporte del machine learning a la inteligencia artificial es la predicción a través de los datos que se tengan almacenados, es decir, los algoritmos de machine learning podrán clasificar, predecir o agrupar datos según las necesidades del problema a resolver, de aquí que su uso ha ido en aumento en los últimos años.

Entre los principales ejemplos de uso que podemos encontrar en los algoritmos de machine learning tenemos la predicción que pueden realizar dentro de vehículos autónomos, de estos se puede mencionar los automóviles de la compañía Tesla, la personalización de contenido según los patrones de comportamiento de un usuario, como se puede apreciar en plataformas de entretenimiento como Netflix, siendo estas, las más populares en el uso de algoritmos de machine learning para recomendar contenido multimedia a los usuarios según el consumo que lleven dentro de la plataforma.

Los algoritmos de machine learning se clasifican en tres grupos: algoritmos supervisados, algoritmos no supervisados y algoritmos de reforzamiento.

2.2.1.1 Algoritmos de machine learning supervisados

Los algoritmos de aprendizaje supervisado, se basan en un conjunto de datos con características (etiquetas) asociadas, se busca que el algoritmo aprenda la relación entre los datos y dichas etiquetas para así aplicar la relación aprendida al momento de recibir nuevos datos nunca antes vistos. Estos algoritmos se pueden clasificar en dos tipos que abordaremos posteriormente: algoritmos de clasificación cuya tarea es asignar una determinada clase a los datos de entrada y los algoritmos de regresión que tienen como objetivo el predecir valores continuos.

El objetivo de los algoritmos de aprendizaje supervisado es obtener modelos de clasificación válidos que permitan trabajar con datos para futuros casos; se busca que el sistema pueda ser capaz de aprender de los datos que ya tiene para poder generalizar y tratar con los datos que aún no tiene, pero en determinado tiempo obtendrá a partir de nueva información (Sierra, 2006).

Un ejemplo común de los algoritmos de aprendizaje automático, es el mostrado en la figura 2.6, donde el servidor de correo electrónico clasifica los nuevos correos electrónicos enviándolos a la carpeta de spam o a la bandeja de entrada basándose en algunas características del usuario, el contenido del correo, el remitente, etc.

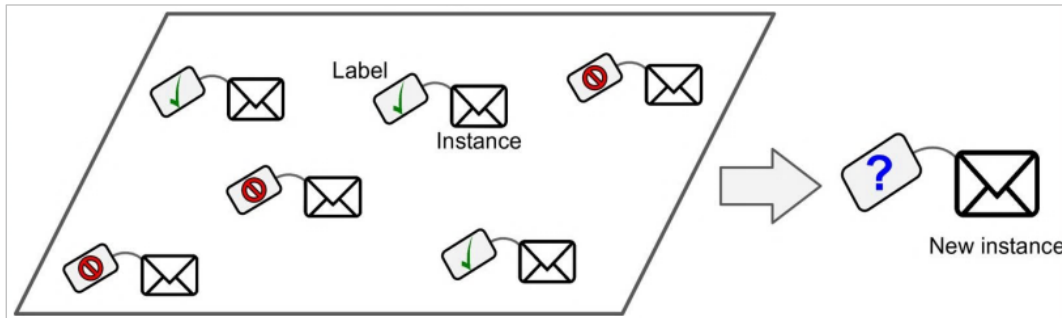


Figura 2.6: Aprendizaje supervisado basado en etiquetas para identificar correo spam

2.2.1.1.1 Algoritmos de clasificación

2.2.1.1.1.1 Vecinos cercanos

El algoritmo K-Nearest Neighbors o vecinos cercanos es un algoritmo fácil de entender, ya que además de ser un algoritmo de alta precisión e insensible a valores atípicos, sus resultados suelen ser altamente competitivos; a menudo, suele usarse como punto de referencia para clasificaciones más complejas como redes neuronales y máquinas de vectores (Harrington, 2012).

Su funcionamiento es simple: se comienza con un conjunto de datos o set de datos de entrenamiento previamente etiquetados, por lo cual es posible saber de antemano las clasificaciones de nuestros datos, entonces, cuando se recibe un nuevo dato sin etiquetar, este se compara con los datos ya etiquetados tomando aquellos datos más parecidos (vecinos cercanos) para revisar sus etiquetas, tomando un valor para k , representado por el número de datos más cercanos al dato sin etiqueta, tal como se ilustra en la figura 2.10, el cual no debería ser mayor a 20, aunque algunos investigadores del tema recomiendan un valor para k en número impar. Una vez asignado el valor para k , se buscarán los k -vecinos más cercanos de la muestra que deseamos clasificar y finalmente se le asignará una etiqueta de la clase por mayoría de votos (Raschka, 2017).

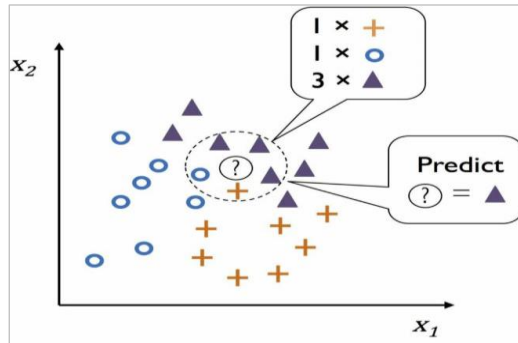


Figura 2.7: Modelo de clasificador k-vecinos cercanos.

2.2.1.1.1.2 Máquinas de soporte vectorial

Según Harrington (2012), las máquinas de soporte de vector son consideradas el mejor algoritmo de clasificación por stock, es decir, entendiendo por stock al hecho de ser un algoritmo que no se modifica desde su base, esto significa que puede tomar el algoritmo desde su base y ejecutarlo, obteniendo una tasa de error muy baja. Es un modelo de aprendizaje muy potente y versátil, capaz de realizar clasificación lineal o no lineal, regresión e incluso detección de valores atípicos (Géron, 2017).

Este algoritmo toma muy buenas decisiones para los puntos de datos que se encuentran fuera del conjunto de entrenamiento manejando un bajo error de generalización, computacionalmente barato y fácil de interpretar sus resultados; aunque suele ser sensible a los parámetros de ajuste y nativamente suele ser un algoritmo de clasificación binaria.

Las SVM (por sus siglas en inglés) se consideran algoritmos de clasificación discriminatorio, dado que los datos de entrenamiento etiquetados, el algoritmo genera un hiperplano óptimo que clasifica los nuevos ejemplos en dos espacios dimensionales. El hiperplano es una línea que divide un plano en dos partes donde cada clase se encuentra en un lado.

Las SVM pueden considerarse una extensión del perceptrón, sin embargo, en este algoritmo el objetivo de la optimización es maximizar el margen, tal como se puede observar en la figura 2.11, es posible observar el margen que se define como la distancia entre el hiperplano de separación (límite de decisión) y el entrenamiento.

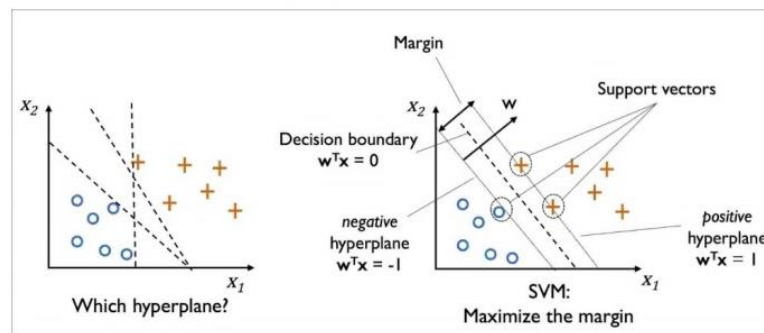


Figura 2.8: Modelo de máquina de soporte vectorial (SVM).

La razón de tener límites de decisión con grandes márgenes es que se tiende a tener un error en la generalización menor mientras que en los modelos con márgenes pequeños se suele ser más propenso a sobrealimentación. Para poder tener una mejor idea de la maximización del margen tomando en cuenta la siguiente ecuación:

$$y^{(i)} \left(w_0 + \mathbf{w}^T \mathbf{x}^{(i)} \right) \geq 1 \quad \forall_i$$

El uso de esta ecuación para obtener el máximo margen suele ser similar al caso de la sigmoide explicada en el algoritmo de regresión logística, ya que buscamos optimizar el problema a través del uso de la función de escalón donde se buscaba llegar a obtener valores cerca de 1 y de 0. En este caso, buscamos trabajar con valores de -1 a 1, ya que solo son diferentes por el signo y la ecuación describe

el margen o que tan cerca se encuentran un punto de datos de nuestro plano de separación y así, no preocuparse si los datos están en la clase -1 o +1 (Harrington, 2012).

2.2.1.1.1.3 Árboles de decisión

Al igual que las SVM (máquinas de soporte vectorial), los árboles de decisión son algoritmos versátiles de aprendizaje automático que pueden realizar tareas de clasificación y regresión. Suelen ser componentes fundamentales de los algoritmos de bosques aleatorios y además no es necesario saber mucho sobre el aprendizaje automático para entender cómo funcionan (Géron, 2017).

Es un algoritmo de supervisión donde dividimos los datos o muestras en dos o más conjuntos homogéneos basados en un diferenciador más significativo en las variables de entrada. Los árboles de decisión se apoyan en técnicas matemáticas y probabilísticas, introduce el concepto de entropía, la cual es una medida de incertidumbre o de desorden, y es usado para ayudar a decidir qué atributo debe ser el siguiente en seleccionarse.

El algoritmo se representa como un grafo en forma de árbol, podemos observar la figura 2.12 donde se encuentran los principales que son el nodo principal o nodo raíz en la parte superior, del cual parten líneas hacia otros nodos inferiores, que a su vez, pueden hacer las veces de nodo raíz; los nodos terminales, son nodos donde termina el flujo y que ya no son raíz de ningún otro nodo, estos, deben tener una respuesta, es decir, la clasificación a que pertenece el objeto sometido en el árbol.

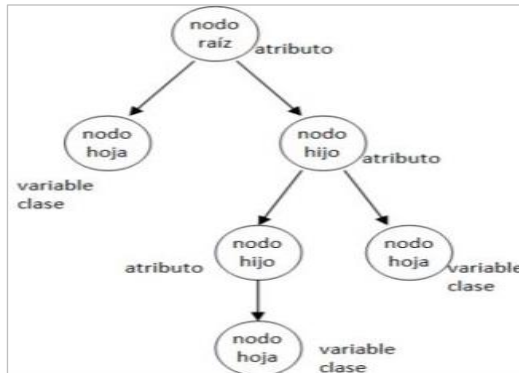


Figura 2.9: Modelo de árboles de decisión.

Es importante dividir el set de datos de manera que nuestros datos no organizados se encuentren más organizados. Una forma de organizar es medir la información utilizando la teoría de la información, la cual se encarga de cuantificar la información, con esto es posible medir de una mejor manera la información, la cual se conoce como entropía de Shannon, proveniente de la teoría de la información de Claude Shannon (Harrington, 2012).

Una vez implementando el árbol es posible evaluar los diferentes criterios de impureza sobre los nodos, por lo cual existen tres métodos a utilizar: entropía, impureza de Gini y el error de clasificación (Raschka, 2017).

En la práctica, tanto la entropía como la impureza de Gini suelen mostrar resultados muy similares, a menudo no vale la pena dedicar mucho tiempo a evaluar arboles de decisión tratando de mezclar ambas técnicas para determinar la impureza. En el caso del error de clasificación, es útil para determinar las ramificaciones del árbol, pero no se recomienda para la toma de decisiones, ya que suele ser menos sensible a los cambios en las probabilidades de las clases de nodos.

2.2.1.1.1.4 Bosques aleatorios

Un algoritmo de bosques aleatorios se puede considerar como un conjunto de árboles de decisión. La idea detrás de este algoritmo es promediar el resultado de múltiples árboles de decisión que lo conforman, ya que individualmente cada árbol posee una alta varianza, permitiendo construir un modelo con mejor rendimiento y menos susceptible a los sobreajustes (Raschka, 2017). En la figura 2.13 se observa un conjunto de árboles de decisión donde los valores de salida de cada elemento suelen ser promediados para obtener un único valor de predicción.

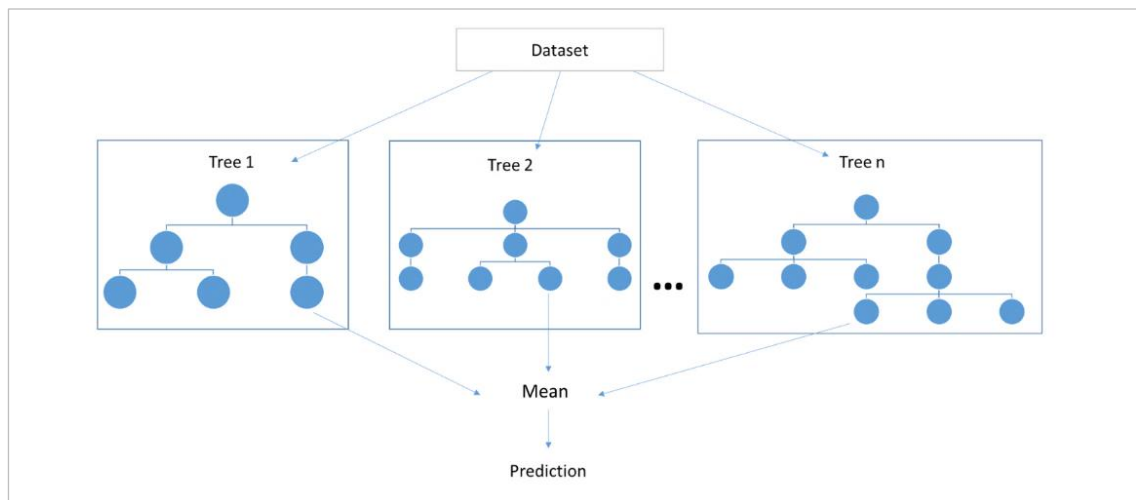


Figura 2.10: Modelo de bosques aleatorios.

Los bosques aleatorios emplean árboles para clasificar un número objeto basado en atributos, donde cada árbol asigna una clasificación y finalmente el bosque elige la clasificación con más votos.

Las principales ventajas que ofrece este algoritmo son:

- Maneja perfectamente grandes cantidades de datos.
- Tiene un método efectivo para estimar datos faltantes y mantiene la precisión cuando falta una gran proporción de datos.

- Suele ser un algoritmo con muy poco control sobre lo que el modelo puede realizar.

2.2.1.2 Algoritmos de machine learning no supervisados

En los algoritmos de aprendizaje supervisado, se sabe de antemano la respuesta que buscamos para nuestro modelo, de aquí partimos en el uso de etiquetas para que nuestro modelo aprenda a clasificar con base en lo que deseamos obtener. En el caso de los algoritmos no supervisados, se trabaja con datos no etiquetados o datos con estructuras desconocidas, con lo cual buscamos que nuestro modelo aprenda sin necesidad de enseñarle a donde deseamos llegar (Géron, 2017).

La principal función de estos algoritmos no supervisados es la de agrupar datos, por lo que el algoritmo debería catalogar los datos de acuerdo a las similitudes que encontrase y así poder crear grupos, es decir, el objetivo de estos algoritmos es pedirle a la máquina que nos diga que tienen en común los datos, para así agruparlos en clústeres (Harrington, 2012).

En el caso del aprendizaje no supervisado no tenemos una variable objetivo como la que se tiene en el aprendizaje supervisado, en lugar de decirle a la máquina que prediga un valor y dado los datos en x , debemos preguntarnos ¿Qué puede decirme la maquina sobre x ?, pudiendo responder preguntas como ¿cuáles son los 3 grupos con más datos agrupados? o ¿cuáles características aparecen juntas con más frecuencia?

2.2.1.2.1 Algoritmo K-means

Uno de los algoritmos más usados en el mundo académico y en la industria, ya que su técnica de agrupación en clúster permite encontrar grupos de objetos similares, objetos que están más relacionados entre sí que con los objetos de otros grupos (Raschka, 2017).

Los ejemplos más comunes de aplicación de agrupación en clúster son las orientadas a negocios que incluyan agrupación de documentos, música y películas de diferentes temas o el encontrar clientes que compartan intereses similares basados en sus comportamientos al comprar. La figura 2.14 muestra el uso de centroides para cada agrupamiento generado,

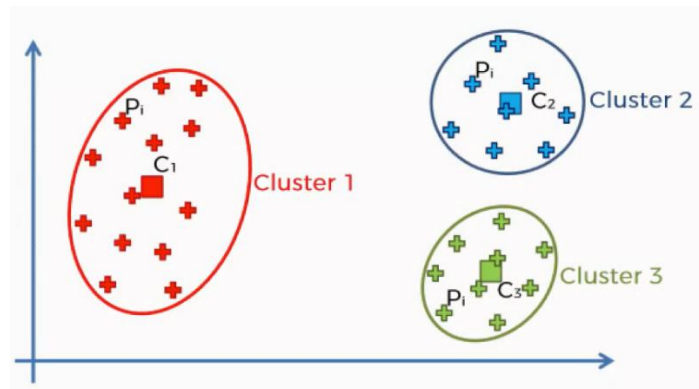


Figura 2.11: Modelo agrupación k-means.

En la figura 2.14, se pueden observar los diferentes agrupamientos basados en la similitud de los datos, es importante saber que el algoritmo k-means pertenece a la categoría de agrupamiento basado en prototipos, es decir, cada agrupamiento está representado por un centroide (promedio) de puntos similares con características continuas, o el medoide (punto más representativo frecuente) en el caso de características categóricas. Si bien este algoritmo es muy bueno en identificar clúster con una forma muy esférica, uno de los inconvenientes es que se debe especificar el número de clúster k a priori, donde una elección inapropiada para k puede resultar en un bajo rendimiento de agrupación.

En las aplicaciones de agrupamiento reales, no se tiene ninguna información de categoría sobre las muestras, en caso contrario, estaríamos hablando de categorías de aprendizaje supervisado, por lo tanto, el objetivo siempre es agrupar las muestras en función de las similitudes de sus características, lo que se logra implementando este algoritmo que puede resumirse en los siguientes cuatro pasos:

1. Seleccionar aleatoriamente k centroides de los puntos de la muestra como centros de agrupación inicial.
2. Asignar cada muestra al centroide más cercano.
3. Mover los centroides al centro de las muestras que se le asignaron.
4. Repetir los pasos 2 y 3 hasta que las asignaciones de clúster no cambien o se alcance una tolerancia definida por el usuario o el número de iteraciones.

En cuanto a la manera de medir la similitud entre los objetos, se define la similitud como el opuesto de la distancia, y una distancia utilizada comúnmente para agrupar con características continuas es la distancia euclidiana cuadrada entre el punto x y el punto y en el espacio dimensional de m .

$$d(\mathbf{x}, \mathbf{y})^2 = \sum_{j=1}^m (x_j - y_j)^2 = \|\mathbf{x} - \mathbf{y}\|_2^2$$

2.2.1.2.2 Agrupamiento jerárquico

El algoritmo de agrupación jerárquica es un enfoque alternativo a la agrupación basada en prototipos, ofreciendo la ventaja de permitir trazar dendogramas (visualizaciones de agrupamiento jerárquico binario), que pueden ayudar a interpretar los resultados mediante la creación de taxonomías significativas. Otra ventaja útil de este enfoque jerárquico es que no es necesario especificar el número de agrupaciones por adelantado (Raschka, 2017).

La agrupación jerárquica se basa en dos enfoques principales que son aglomerados y divididos al momento de agrupar. En la primera, se comienza con una agrupación que abarca todas las muestras dividiendo iterativamente la agrupación en minúsculos hasta que cada agrupación solo contenga una

muestra; en el segundo caso, se comienza cada muestra como un clúster individual y combina los pares de clústeres más cercanos hasta que solo quede un clúster. En la figura 2.15 podemos observar un modelo de agrupamiento jerárquico implementando dendogramas para llegar a un solo agrupamiento.

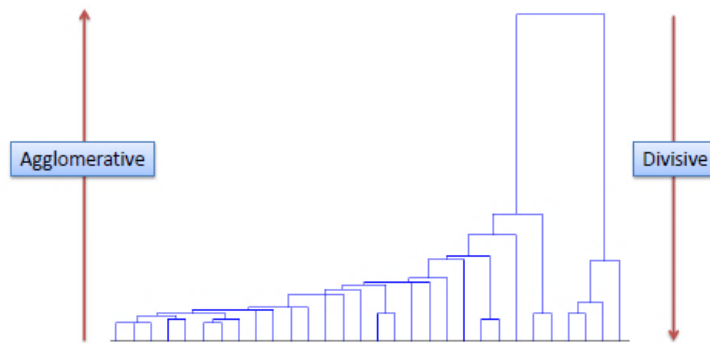


Figura 2.12: Modelo agrupación jerárquico.

El agrupamiento jerárquico es un procedimiento iterativo que puede resumirse mediante los siguientes pasos:

1. Calcular la matriz de distancia de todas las muestras.
2. Representa cada punto de datos como un clúster único.
3. Combinar los dos grupos más cercanos según la distancia entre los miembros más distantes.
4. Actualizar la matriz de similitud.
5. Repetir los pasos del 2 al 4 hasta que quede un solo grupo.

2.3 Sistemas web

Un sistema según Senn (1995), es un conjunto de componentes que interactúan entre sí para lograr un objetivo en común, de la definición anterior podemos decir, que un sistema web es aquel conjunto de herramientas disponibles en un servidor web donde los usuarios pueden acceder a ellas a través del navegador web y una conexión disponible de internet.

Los sistemas web se construyen utilizando un lenguaje de programación que sea capaz de interpretar el navegador web. Las aplicaciones web, como también se les conoce a los sistemas web, son populares debido a la practicidad al no tener que instalar programas para su uso, como lo suelen ser las conocidas aplicaciones de escritorio, además, hoy en día, el uso de aplicaciones en un entorno web suele ser muy populares por debajo de las aplicaciones móviles.

2.3.1 Elementos de un sistema web

Dentro de la arquitectura de software es posible encontrar diversos componentes que permiten generar un ambiente adecuado para el desarrollo del software. En un entorno web, es posible determinar varios elementos que interactúan entre sí, compuesta principalmente por cliente, servidor, bases de datos, back end y front end.

2.3.2 Definición de cliente-servidor

Un servidor es una computadora conectada a la red que ofrece un servicio a los clientes a través de peticiones, donde su función es almacenar archivos y servir de aquella información que solicita el cliente (Castedo et al., n.d., 2008), donde algunos servidores pueden ser desde una computadora sencilla hasta computadoras muy potentes. Ejemplos de servidores podemos encontrar aquellos enfocados a ofrecer servicios en la nube, correo electrónico, bases de datos, DNS, entre otros.

Por otra parte, el cliente suele ser aquel usuario que interactúa con el servidor a través de una interfaz, que suele denominarse front end, donde la interacción está basada en enviar peticiones y recibir respuesta por parte del servidor. La figura 2.17, muestra cómo se da el flujo de trabajo dentro de la arquitectura cliente-servidor, como se explicó anteriormente, un cliente solicita información mediante peticiones al servidor, este responde al cliente mediante respuestas, todo lo anterior utilizando mayormente el protocolo HTTP.

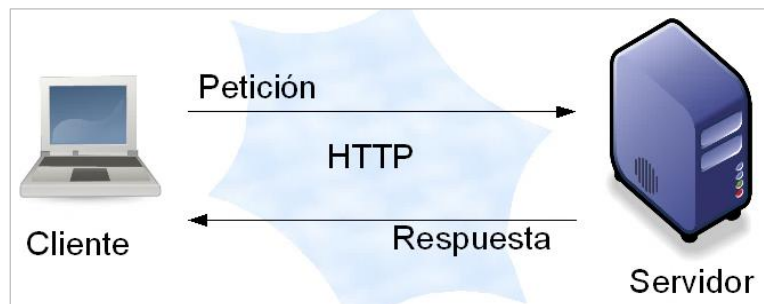


Figura 2.13: Arquitectura cliente-servidor.

2.3.3 Definición de Front End y Back End

En la arquitectura de software se le conoce al Front End como la interfaz sobre la cual interactúan los usuarios de un determinado sistema y es responsable de recolectar los datos y prepararlos para enviarlos al servidor, dicho de otra manera, el Front End es el conjunto de vistas representados en la interfaz GUI que busca la interacción con el usuario a través de un ordenador mediante el uso de formularios y otras ventanas emergentes es la parte del servidor (Gunnulfsen, 2013).

Por otra parte, el Back End es el motor encargado de procesar la información enviada desde el Front End. Es considerada la parte lógica del sistema y aunque no es visible por el cliente, ofrece las funcionalidades necesarias para mantener el sistema estable. El Back End trabaja directamente con

los lenguajes de programación, optimizando los recursos, revisando la seguridad y manteniendo los estándares de accesibilidad (Garfias, 2012).

En la figura 2.18, se puede observar la representación del front end y back end, mientras el front end se encarga de ofrecer las vistas y navegabilidad al usuario, es posible representarse a través de maquetación HTML, hojas de estilo CSS y contenido dinámico con JavaScript que ofrecen una mejor experiencia de uso para el usuario, mientras que en el lado del Back End, se puede encontrar la lógica del negocio desarrollada con algún lenguaje de programación de alto nivel que permita mantener la estabilidad y seguridad del sistema.

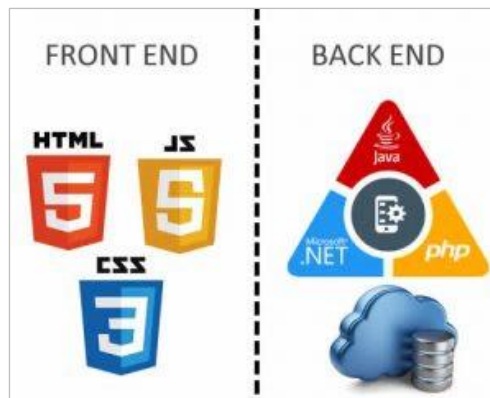


Figura 2.14: Ejemplificación del front end y back end.

2.4 Trabajos relacionados

Con la finalidad de abordar las soluciones propuestas para diseñar e implementar un algoritmo de aprendizaje automático que ayude en los procesos de alfabetización del niño, se presentan dos proyectos que comparten cierta relación sobre todo en el apoyo de regularizar niños con problemas de aprendizaje.

2.4.1 Deslixate

Deslixate es una aplicación móvil creada por Sandra Karen Cadena, licenciada en Pedagogía y Cesar López Martínez, ingeniero en computación de la Facultad de Estudios Superiores Aragón de la UNAM. La aplicación alerta si existen indicios de dislexia fonológica, dislexia profunda o dislexia superficial en niños entre siete y doce años, buscando que los niños que la utilicen se sientan libres del estrés generado debido a sus problemas y dificultades escolares a la vez que trabajan en ejercicios para regularizarse.

La aplicación cuenta con 10 ejercicios presentados mediante juegos que buscan evaluar los tres tipos de dislexia mencionados en un tiempo aproximado de 25 minutos, al término, el programa muestra un pre diagnóstico que indica si hay algún riesgo de dislexia (Méndez, 2018).

La aplicación tiene cinco opciones para comenzar a trabajar donde es necesario registrar un usuario para comenzar con el test, además, ofrece la opción de actividades individuales, que son para reforzar cada uno de los tipos de dislexia, permitiendo que los usuarios puedan interactuar y aprender de una manera divertida.

Deslixate ofrece un complemento llamado semáforo, el cual indica con los colores del mismo, cómo es el avance del niño, en el caso del color verde, todo se encuentra bien, para el color amarillo, señala algunos fallos en reactivos sin ser nada grave, y el color rojo indica que la actividad se hizo mal, tomándola como una alerta. Al finalizar el test, es posible guardar los resultados en formato PDF con los datos del diagnóstico, el número de respuestas correctas e incorrectas y el tiempo empleado, tal como se muestra en la figura 2.19.

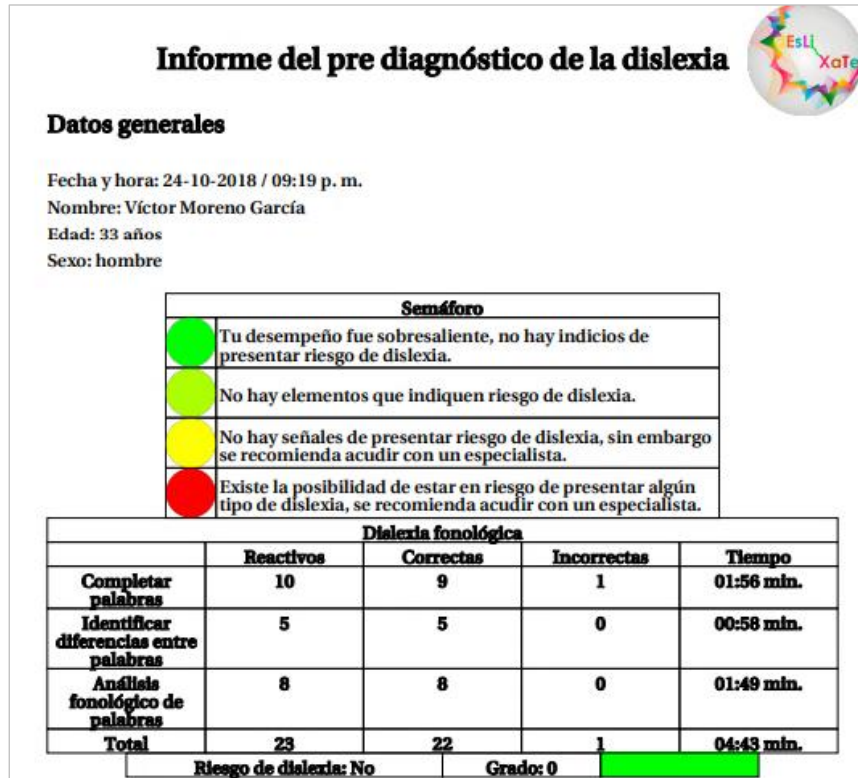


Figura 2.15: Informe de usuario ofrecido por la aplicación Deslixate

La aplicación se encuentra desarrollada para funcionar solamente en dispositivos Android, se puede descargar desde la Play Store de Android y actualmente se encuentra en su versión 2.0.8 con actualización en mayo del 2019 y más de 1000 descargas. Aunque no se menciona en sus reseñas, la aplicación no implementa algún tipo de algoritmo de aprendizaje automático, ya que los ejercicios ya se encuentran definidos y el objetivo de la aplicación es solo pre diagnosticar en base a los resultados y tiempos durante la evaluación.

2.4.2 DyTECTIVE

DyTECTIVE es una herramienta para la detección y mejora de las habilidades relacionadas con la dislexia la cual puede ser utilizada desde niños de 7 años hasta adultos. DyTECTIVE ofrece un test de

15 minutos mediante una serie de pruebas acordes a la edad, con el objetivo de detectar si existe algún riesgo de tener dislexia, además, ofrece actividades de lectura y escritura a través de 42000 juegos que se personalizan en función de 24 habilidades cognitivas.

La aplicación ofrece la posibilidad de comenzar realizando un test de cribado, cuyo resultado se consideraría un pre diagnóstico, además, es posible realizar actividades de juego, pero con una opción de pago, aunque los nuevos usuarios cuentan con siete días de prueba DyetectiveU, un juego interactivo con misiones donde es posible realizar actividades de competencia lingüística, trabajo de memoria, procesos perceptivos visuales, funciones ejecutivas y procesos perceptivos audiovisuales.

El éxito de la aplicación se debe a la implementación de un algoritmo de aprendizaje automático, ya desde el 2015 la Dra. Luz Rello, quien dirige este proyecto, viene trabajando con algoritmos de aprendizaje automático para la detección de la dislexia a partir de datos derivados en el uso de un eye-tracking provenientes de experimentos realizados entre los años 2010 y 2014, entrenando un modelo con alrededor de 1135 lecturas registradas, pertenecientes a 97 personas y casi la misma cantidad con y sin dislexia, cuya lengua es el español (Rello, 2018).

El modelo utiliza un clasificador binario lineal basado en máquinas de soporte vectorial (SVM) y alcanza un 80% de exactitud con las variables más informativas. Según la Dra. Rello, las variables que le resultaron más útiles para la clasificación fueron el tiempo de lectura, la duración media de las fijaciones y la edad del participante. Se crearon alrededor de 212 ítems, agrupados en 32 conjunto de ejercicios compuestos principalmente de ejercicios lingüísticos y atencionales contemplando las siguientes competencias lingüísticas: alfabética, conciencia fonológica, silábica, léxica, morfológica, sintáctica, semántica y ortográfica, así como otros factores relacionados con la dislexia, entre ellos

procesos perceptivos de discriminación y categorización visual (Rello L., 2018). La figura 2.20 ilustra una vista de un ejercicio a desarrollar dentro de la aplicación.



Figura 2.16: Ejercicio dentro de la aplicación Dyetective

Dyetective se encuentra diseñado para funcionar en dispositivos Android y IOS, es posible descargarla desde la Play Store para Android en su versión gratuita, ya que el poder disponer de la versión con los 42000 ejercicios, tiene un costo que debe ser cubierto. Dyetective se encuentra actualmente en su versión 1.5.14 con actualización en abril del 2019 y más de 10000 descargas a nivel de habla hispana.

Capítulo 3

Análisis y diseño

Este capítulo presenta el modelo de algoritmo propuesto, así como el análisis y diseño a implementar. Para llevar a cabo lo anterior, se utiliza una metodología la cual permitirá analizar el problema y poder describir las necesidades del algoritmo y su respectivo sistema web a implementar desde diferentes escenarios.

Para poder desarrollar el algoritmo, se siguió una metodología de tres etapas: la primera etapa consiste en el estudio del estado del arte, permitiendo determinar los alcances y limitaciones, a través de un proceso detallado de recopilación de información de los temas relacionados, buscando identificar las posibles soluciones.

En la segunda etapa, se incluye el análisis y diseño del algoritmo, que inicia con la información recopilada en la etapa anterior para así obtener la arquitectura del sistema web propuesto. Finalmente, la tercera etapa, está formada por la implementación del algoritmo en el sistema web, además del correspondiente análisis de resultados, que será de utilidad para verificar el correcto funcionamiento a través de las pruebas de usabilidad.

3.1 Análisis del sistema

El sistema web busca ayudar a niños con problemas de dificultades específicas del aprendizaje, ofreciendo una determinada cantidad de ejercicios para resolver a través de 4 categorías según los formatos de alfabetización utilizados por la Secretaría de Educación y Cultura de Sonora, la cual clasifica a los niños durante su proceso de alfabetización en pre silábico, silábico, silábico alfabético y alfabético de acuerdo al desempeño registrado por el maestro en el aula de clases y el desarrollo del niño.

El sistema tiene como objetivo inicial el asignar una categoría al alumno cuando se capturen los valores de su alfabetización, dicha asignación corresponderá a la predicción realizada por el algoritmo de machine learning puesto ya en producción, el cual va a trabajar con diecinueve características que forman parte de dicho formato, en donde se busca identificar entre discapacidades, dificultades severas, trastornos y aptitudes sobresalientes de cada uno de los alumnos pertenecientes al sistema de educación básica del estado de Sonora. Con las características anteriores, el algoritmo entrenado podrá ser capaz de asignar una de las cuatro categorías de ejercicios donde el alumno será capaz de reforzar con el conocimiento aportado por el docente en el aula de clase. Además, una vez que el alumno sea clasificado por el algoritmo, tendrá a su disposición un sistema web para el desarrollo de sus actividades y donde el maestro, podrá observar el avance de las actividades de cada uno de sus alumnos registrados.

El docente, por su parte, tendrá a su disposición un área para la administración de sus grupos, alumnos, generación de reportes para conocer el avance y desempeño de sus alumnos.

3.1.1 Arquitectura del sistema

A continuación, se observa en la figura 3.1 la arquitectura general del sistema, donde es posible observar los componentes que la integran: el sistema web que ofrece las vistas necesarias para el trabajo, entre otros módulos que se presentan a continuación:

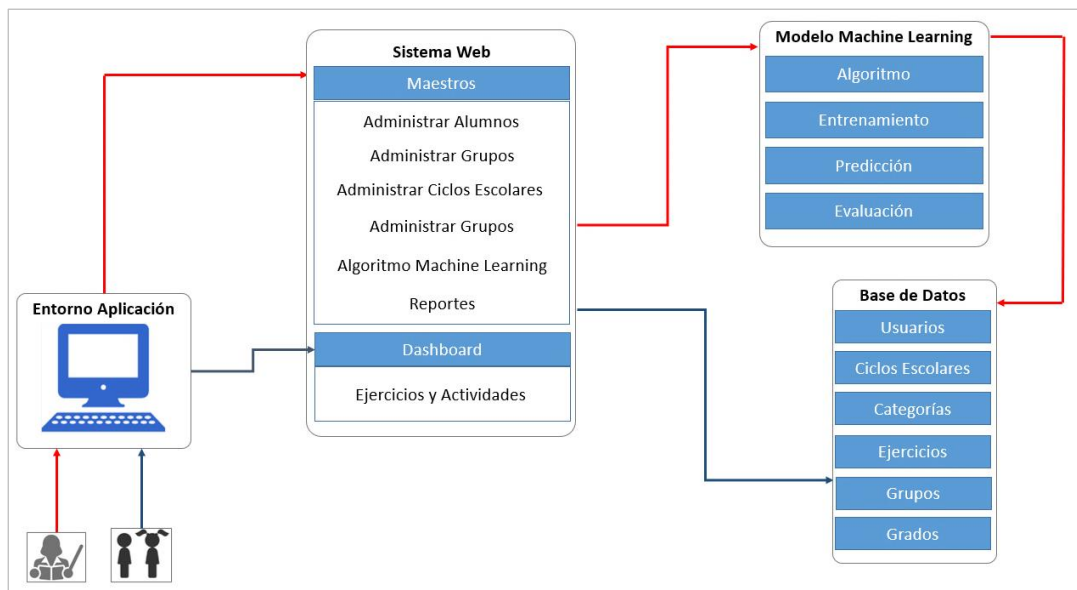


Figura 3.1: Arquitectura del sistema web con implementación de algoritmo de machine learning

El módulo de sistema web, ofrece las áreas de trabajo para alumnos y maestros, que permite a estos últimos, administrar las funciones básicas del sistema web crear, leer, actualizar y borrar (CRUD, por sus siglas en inglés) de grupos y alumnos, además, de contar con un área de generación de reportes para ver el desempeño del alumno.

El módulo de machine learning es el principal eje del proyecto, ya que es aquí donde se llevará a cabo la implementación del algoritmo de aprendizaje automático que asignará la categoría sobre la cual trabajará cada uno de los alumnos del sistema. Su importancia radica en el entrenamiento sobre el cual aprendió la categorización basándose en los procesos de alfabetización, permitiendo evaluar

las características de cada uno de los alumnos registrados, siendo capaz de predecir las cuatro categorías ya mencionadas anteriormente.

3.1.2 Diagramas de contexto nivel 0

Es importante que al desarrollar un sistema se definan de manera legible los diferentes modelos que formarán parte del mismo, permitiendo comprender de una manera gráfica el funcionamiento del sistema. La figura 3.2 representa el diagrama de contexto nivel 0, el cual, considera tres entidades que interactúan con el sistema web, siendo los alumnos, el que solicita las actividades a desarrollar, el maestro, que administra grupos, alumnos y genera reportes; y finalmente, el algoritmo, encargado de asignar una determinada categoría a cada alumno registrado dentro del sistema.

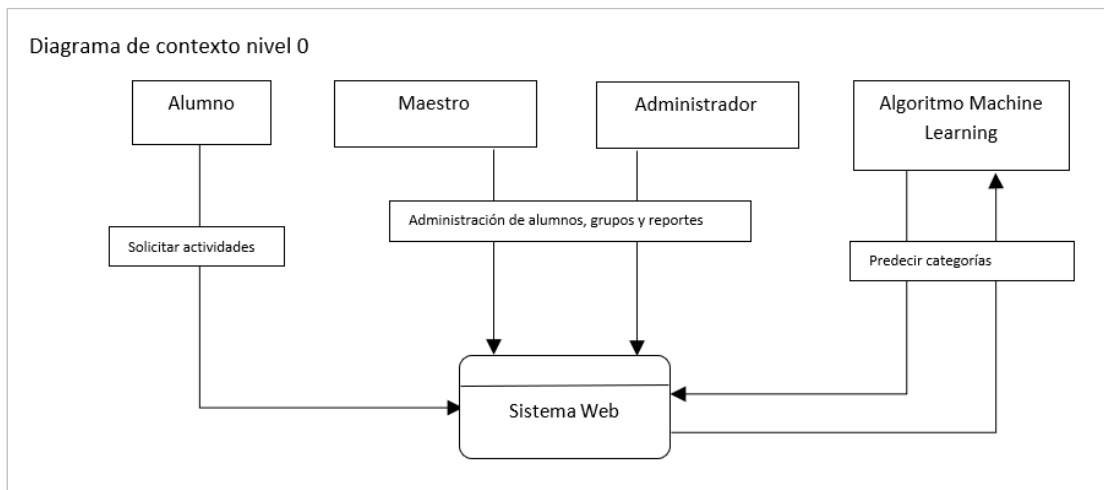


Figura 3.2: Diagrama de contexto nivel 0 para sistema web

3.1.3 Diagramas de contexto nivel 1

El proceso visto en el diagrama anterior, se puede desglosar mediante un diagrama de contexto nivel 1, que nos ayuda a mostrar un mayor detalle el funcionamiento del sistema web, y en la figura 3.3, es posible observar el diagrama con sus respectivas entidades y procesos que lo conforman.

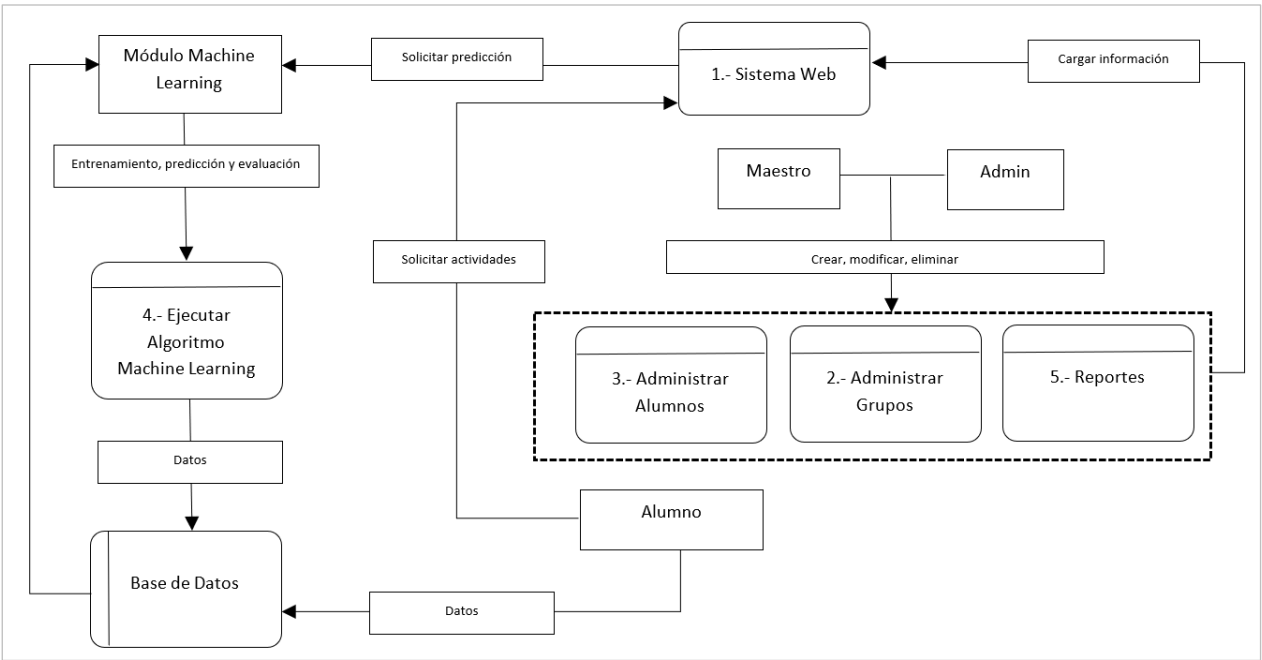


Figura 3.3: Diagrama de contexto nivel 1 para sistema web

Como se puede observar en la figura anterior, al detallar los procesos podemos encontrar cómo se va dando el flujo de actividades dentro del sistema. Los procesos suelen ser llamados por el maestro, ya que es importante administrar sus alumnos y grupos, para poder llamar al proceso que emplea el algoritmo de machine learning y se asignen los ejercicios al alumno.

3.1.4 Diagramas de casos de uso

En esta sección se identifican las necesidades que el sistema debe cubrir, utilizando el lenguaje de modelado unificado (UML), se utilizan casos de uso para describir de manera general las actividades que realiza el sistema.

Los casos de uso del presente proyecto proveen una breve descripción de las interacciones que pueden llevarse a cabo entre los usuarios y el sistema, identificando seis casos de uso: inicio de sesión,

ejercicios, administración de alumnos, administración de grupos, reportes y módulo de machine learning, que son llevadas a cabo por los usuarios alumno y maestro.

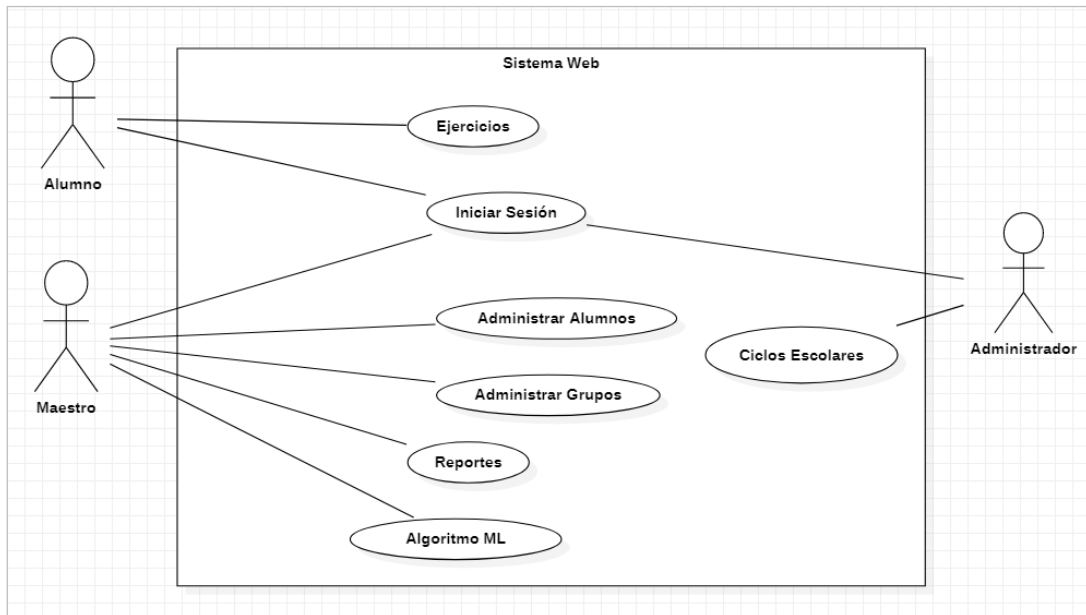


Figura 3.4: Diagrama de casos de uso para sistema web

En esta sección se identifican las necesidades que el sistema debe cubrir, utilizando el lenguaje de modelado unificado (UML), se utilizan casos de uso que son empleados para describir de manera general las actividades que realiza el sistema. A continuación, se presentan las definiciones correspondientes para los casos de usos de la figura 3.4:

Tabla 3.1: Caso de uso: iniciar sesión

Nombre: Iniciar sesión.
Descripción: Permite tener acceso al sistema
Actores: Usuario no logueado.

<p>Precondiciones: El usuario debe tener conexión a internet.</p>
<p>Flujo Normal:</p> <ol style="list-style-type: none"> 1. El actor escribe el nombre de usuario / email en el campo usuario. 2. El actor escribe la contraseña en el campo password 3. El actor pulsa el botón iniciar sesión 4. El sistema comprueba la validez de los datos y su rol dentro del sistema. 5. El sistema carga la ventana de sesión correspondiente.
<p>Flujo Alternativo:</p> <ol style="list-style-type: none"> 1. El sistema comprueba la validez de los datos, si los datos no son correctos, se avisa al actor de ello permitiéndole que los corrija.

Tabla 3.2: Caso de uso: ejercicios

<p>Nombre: Ejercicios</p>
<p>Descripción: Permite al actor realizar sus actividades correspondientes.</p>
<p>Actores: Usuario alumno</p>
<p>Precondiciones: El usuario debe estar logueado</p>
<p>Flujo Normal:</p> <ol style="list-style-type: none"> 1. El actor pulsa sobre el botón de ver avance, el cual le mostrará las actividades que ya ha realizado y su calificación 2. El actor puede elegir realizar actividades ya elaboradas, o trabajar con la actividad más reciente pulsando el botón de comenzar actividad. 3. El actor desarrolla actividad. 4. Al término de la actividad, el actor debe pulsar botón de guardar resultados 5. El sistema almacena la calificación obtenida del autor y actualiza dashboard
<p>Flujo Alternativo:</p>

1. El sistema le ofrece la posibilidad de iniciar actividades, ya que no encontró registros de calificaciones.

Tabla 3.3: Caso de uso: administrar grupos

<p>Nombre: Administrar grupos</p>
<p>Descripción: Permite al actor realizar operaciones CRUD a los grupos de alumnos.</p>
<p>Actores: Usuario maestro, usuario administrador</p>
<p>Precondiciones: El usuario debe estar logueado, el usuario debe tener creado un grupo.</p>
<p>Flujo Normal:</p> <ol style="list-style-type: none"> 1. El actor puede pulsar el botón agregar alumno. 2. El actor registra los datos del alumno 3. El actor selecciona el grupo asignado para el alumno. 4. El actor pulsa sobre el botón guardar alumno. 5. El sistema comprueba la validez de los datos. 6. El sistema envía mensaje satisfactorio.
<p>Flujo Alternativo:</p> <ol style="list-style-type: none"> 1. El actor puede pulsar sobre el botón importar alumnos. 2. El actor selecciona el archivo en formato CSV con las características adecuadas para la correcta importación. 3. El actor selecciona el grupo al cual se importará los alumnos. 4. El actor pulsa sobre el botón importar alumnos. 5. El sistema comprueba la validez de los datos. 6. El sistema envía mensaje satisfactorio.

Tabla 3.4: Caso de uso: administrar alumnos

Nombre: Administrar alumnos.
Descripción: Permite al actor realizar operaciones CRUD de sus alumnos.
Actores: Usuario maestro y usuario administrador
Precondiciones: El usuario debe estar logueado, el alumno debe estar ya registrado.
Flujo Normal: <ol style="list-style-type: none"> 1. El actor puede pulsar el botón ver grupos. 2. El actor selecciona la opción detalles de grupo. 3. El actor pulsa el botón agregar, modificar o alumno. 4. El actor actualiza la información del alumno. 5. El sistema comprueba la validez de los datos. 6. El sistema envía mensaje satisfactorio.
Flujo Alternativo: <ol style="list-style-type: none"> 1. El sistema comprueba la validez de los datos, si los datos no son correctos, se avisa al actor de ello permitiéndole que los corrija.

Tabla 3.5: Caso de uso: reportes

Nombre: Reportes
Descripción: Permite al actor ver reportes de aprovechamiento de sus alumnos.
Actores: Usuario maestro y usuario administrador
Precondiciones: El usuario debe estar logueado.
Flujo Normal:

<ol style="list-style-type: none"> 1. El actor puede pulsar el botón reportes 2. El actor selecciona el tipo de reporte a visualizar 3. El sistema comprueba la validez de los datos. 4. El sistema muestra del reporte
<p>Flujo Alternativo:</p> <ol style="list-style-type: none"> 1. El sistema comprueba la validez de los datos, si los datos no son correctos, se avisa al actor de ello permitiéndole que los corrija.

Tabla 3.6: Caso de uso: modulo machine learning

<p>Nombre: Módulo machine learning.</p>
<p>Descripción: Permite al actor asignar una categoría al alumno mediante el uso del algoritmo de machine learning</p>
<p>Actores: Usuario maestro y usuario administrador</p>
<p>Precondiciones: El usuario debe estar logueado, el alumno debe estar ya registrado.</p>
<p>Flujo Normal:</p> <ol style="list-style-type: none"> 7. El actor puede pulsar el botón predecir categoría. 8. El actor selecciona la opción detalles de grupo. 9. El actor selecciona a los alumnos. 10. El actor pulsa el botón predecir categoría 11. El sistema comprueba la validez de los datos. 12. El sistema llama al módulo de machine learning. 13. El módulo revisa el set de datos para comprobar su existencia. 14. El módulo realiza un entrenamiento con los datos almacenados. 15. El módulo realiza una predicción con los datos registrados del alumno. 16. El sistema recibe la categoría y procede a la asignación de ejercicios.

Flujo Alternativo:

2. El sistema comprueba la validez de los datos, si los datos no son correctos, se avisa al actor de ello permitiéndole que los corrija.

3.2 Diseño del sistema

En esta sección se muestra el diseño del sistema web y del algoritmo de aprendizaje automático que se va a utilizar para la predicción de categorías, además, se muestran los diagramas que dan explicación al sistema a través de las diferentes perspectivas. Se ha optado por tomar algunos diagramas basados en el modelo de 4 + 1 vistas, lo anterior, debido a que suelen ejemplificar de una manera muy practica el funcionamiento y modularidad del sistema.

3.2.1 Diagramas conceptual de clases

Una vez que ya se tiene identificado los requerimientos del sistema mediante los casos de uso, se realiza un diagrama conceptual de clases, que permite la comprensión del contexto del problema y expresarlo a través del paradigma de programación orientada a objetivos, sirviendo de base para el desarrollo del sistema web y del algoritmo de machine learning, tal como se muestra en la siguiente figura 3.5.

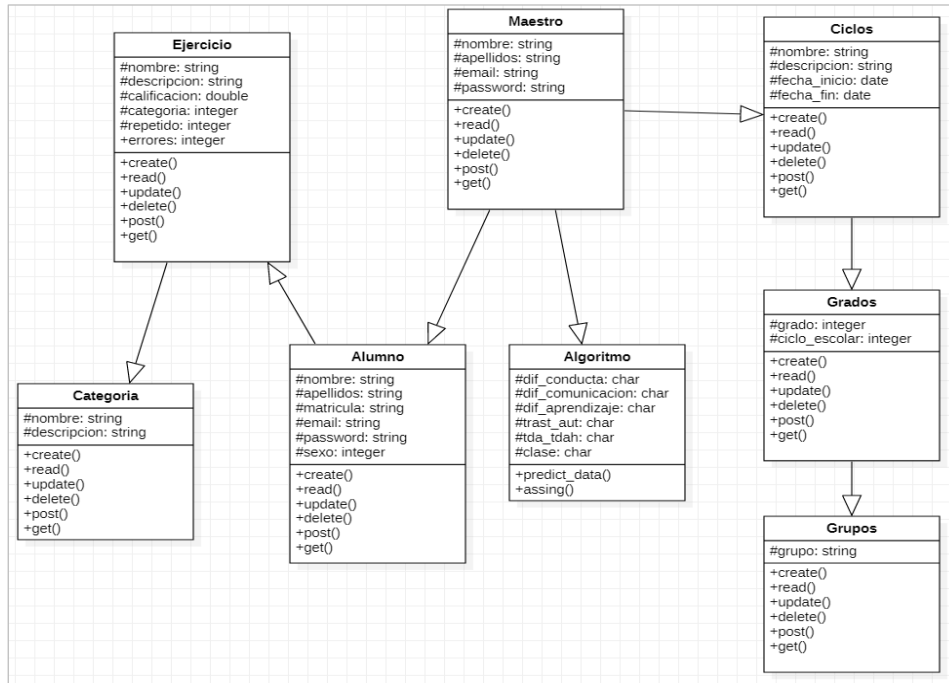


Figura 3.5: Diagrama de clases para sistema web

3.2.2 Diagrama de secuencia

Un diagrama de secuencia nos muestra la interacción que existe entre un conjunto de objetos de un sistema en función del tiempo en que se está utilizando, indicando los módulos y clases que formarán parte del sistema y las respectivas llamadas a cada una de las funciones de las clases. Observemos la figura 3.6, que nos ilustra el diagrama de secuencia propuesto para la asignación y predicción de categorías de un determinado alumno, actividad que realiza el maestro solicitando al algoritmo de machine learning ya entrenado predecir nueva información.

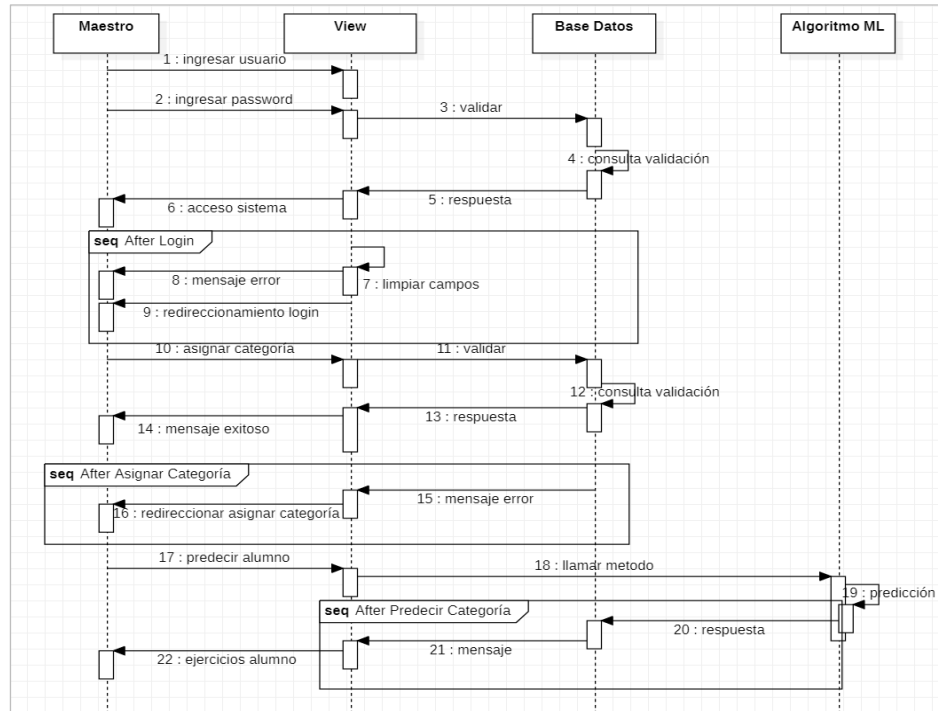


Figura 3.6 Diagrama de secuencia de asignación de categoría a un alumno a través del algoritmo de machine learning

Como se observó la figura 3.6, es posible seguir la secuencia que tanto el usuario maestro como el algoritmo deben llevar para realizar la predicción de categorías mediante el ingreso de nueva información del usuario alumno, podemos ver en la secuencia la validación de la información, es decir, el sistema web debe revisar que exista la información necesaria del alumno para poder realizar una predicción, de lo contrario, será imposible asignarle una categoría con ejercicios al alumno.

A continuación, se presenta la figura 3.7 que muestra el diagrama de secuencia donde el alumno interactúa con el sistema para solicitar ejercicios según su categoría previamente asignada por el algoritmo de machine learning, donde solicita al sistema le muestre los diferentes ejercicios a responder.

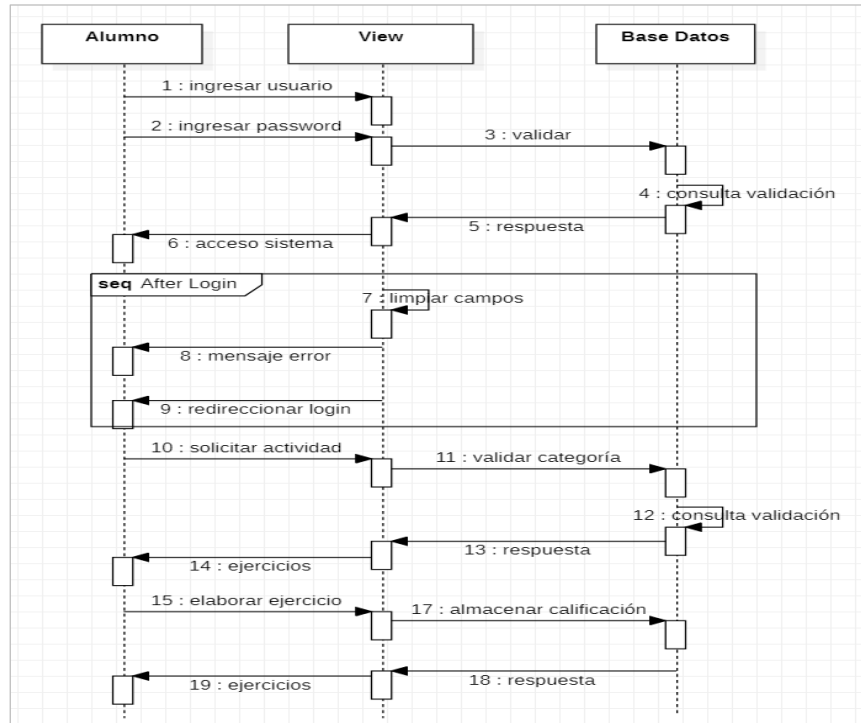


Figura 3.7 Diagrama de secuencia de elaboración de ejercicios por parte del alumno

3.2.3 Diagramas de componentes y despliegue

Un diagrama de componentes y despliegue es útil para conocer la distribución física de los componentes de software ofreciéndonos una visión más general del despliegue de nuestro sistema web, se conforma principalmente de nodos y conectores, ofreciéndonos la relación existente entre los diferentes elementos desplegados, así como los componentes lógicos de los distintos nodos.

En la figura 3.8, se observa de manera combinada los elementos que integran nuestro diagrama, donde podemos observar dos nodos, uno que corresponde al servidor, donde se encuentran los componentes base de datos, el framework laravel y la API del algoritmo que son los elementos mínimos para dar el correcto funcionamiento al sistema. Por otro lado, encontramos la aplicación web donde encontramos la estructura conformada por los componentes modelos, que permiten la conexión

a la base de datos, los controladores, que en nuestro framework ofrecen la lógica del sistema y finalmente las vistas, que son las diferentes páginas sobre las cuales el usuario final va a interactuar.

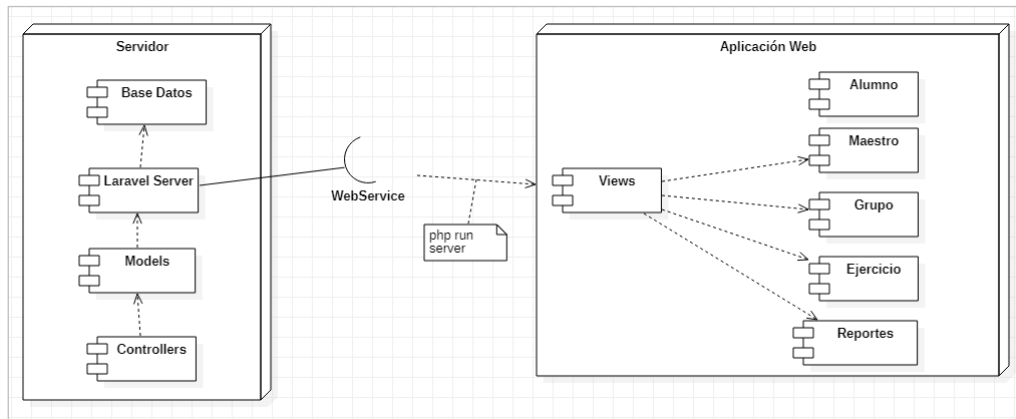


Figura 3.8 Diagrama de componentes y despliegue para el sistema web

3.2.4 Diagrama de actividades

Un diagrama de actividades ayuda a conocer el flujo de actividades que son posibles realizar dentro del sistema, ya que a través del diagrama se puede visualizar el trabajo desde el punto inicial hasta su punto final de una manera más sencilla de comprender detallando las posibles rutas que existen durante el desarrollo de eventos.

La figura 3.9 presenta el diagrama de actividades que realiza el maestro para poder dar de alta alumnos dentro del sistema web, se observan las diferentes validaciones realizadas con la base de datos para evitar la duplicidad de información, así como el flujo llevado a cabo en caso de realizar el procedimiento de una manera correcta.

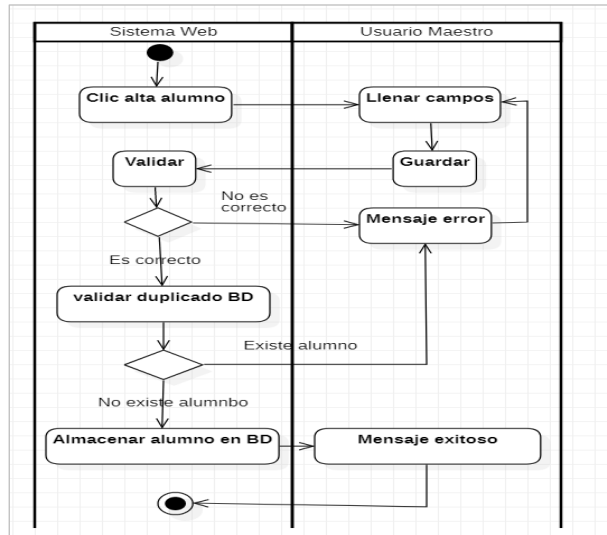


Figura 3.9 Diagrama de actividades para el registro de alumnos en el sistema web

La figura 3.10 presenta el diagrama de actividades que realiza el alumno para resolver un determinado ejercicio previamente categorizado por el algoritmo de machine learning, se puede observar el flujo llevado a cabo al iniciar sesión, solicitar el ejercicio, resolverlo y que este sea calificado por el sistema y posteriormente almacenado en sus registros dentro de la base de datos.

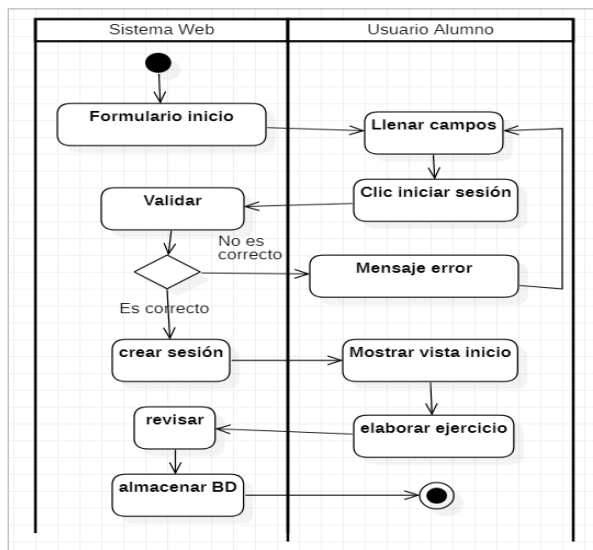


Figura 3.10 Diagrama de actividades para elaboración de ejercicio del alumno en el sistema web

Finalmente, la figura 3.11 muestra el flujo de actividad realizado por el algoritmo al momento de realizar una predicción, como se había mencionado anteriormente, el usuario maestro solicita la predicción dentro del sistema, seleccionando al alumno correspondiente, es aquí donde el sistema verifica si el alumno ya tiene una categoría asignada, en caso contrario, llama al algoritmo para realizar una predicción sobre la información capturada durante el proceso de registro, al finalizar, el algoritmo envía la predicción y esta se almacena en los registros de la base de datos correspondientes al alumno.

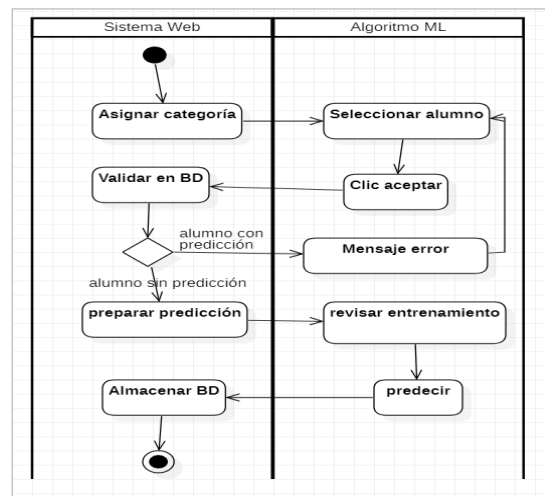


Figura 3.11 Diagrama de actividades para la asignación de una categoría a un determinado alumno registrado

3.2.5 Esquema de la base de datos

A continuación, se presenta el esquema de la base de datos que será utilizada para almacenar la información del sistema web y el set de datos utilizado para el entrenamiento del algoritmo de machine learning, la figura 3.12 ilustra el modelo relacional de la base de datos del sistema.

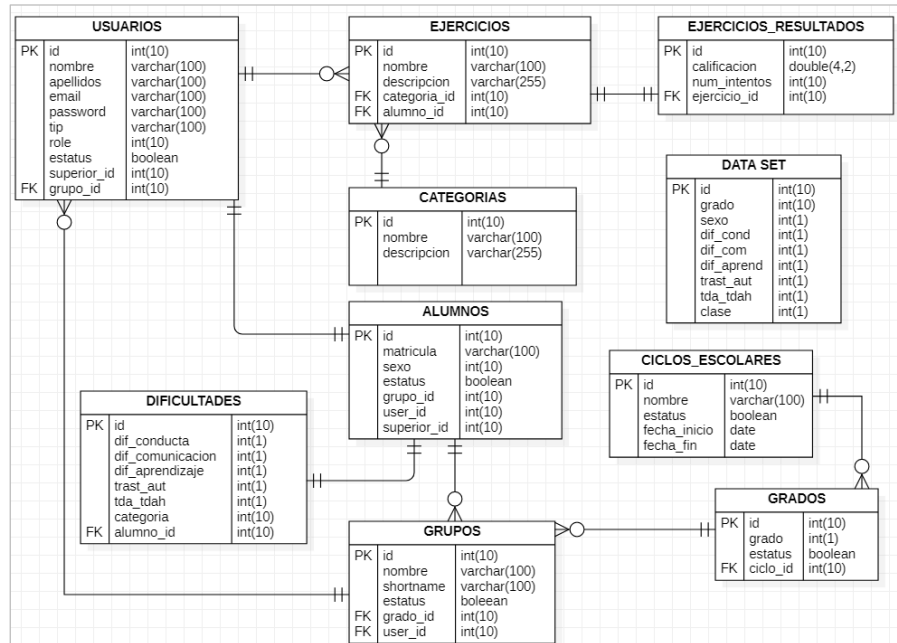


Figura 3.12 Esquema relacional de la base de datos para el sistema web

Como se observa en la figura anterior, el diagrama de la base de datos muestra diez tablas que ayudarán a la correcta administración de la información, los usuarios registrado, la tabla de ciclos escolares donde se encuentran relacionadas grados y grupos y finalmente, la tabla de categorías con su respectiva relación con la tabla de ejercicios y una tabla donde se almacenarán las calificaciones de los ejercicios. Además, se observa una tabla llamada data set, la cual contendrá la información de datos que utilizará el algoritmo para su entrenamiento y predicción.

3.3 Proceso de selección de modelo de machine learning

Cuando ya se han realizado los dos pasos anteriores y ya se ha podido analizar la matriz de confusión, obteniendo buenos resultados en su diagonal, es posible realizar predicciones con nuevos

datos y seleccionar el algoritmo que se considere mejor su desempeño durante el proceso de selección, en términos generales, el algoritmo se encuentra listo para su implementación.

En la figura 3.13 se pueden observar las fases completas de identificación y selección de un algoritmo de machine learning, tomando en cuenta lo visto en este último tema, el diagrama explica de una forma muy práctica todo el procedimiento que se llevó a cabo para poder seleccionar adecuadamente un algoritmo para el sistema web, y como se observa, se inicia con la recolección de la información para integrar un set de datos que posteriormente pueda ser normalizado y se identifiquen los valores que formarán parte del vector de características. Después, es necesario realizar determinados entrenamientos para reconocer que algoritmo tiene un mejor desempeño sobre los datos, incluso, en ocasiones es necesario realizar reestructuraciones y modificaciones al set de datos, esto a través de la evaluación de los modelos para conocer si será necesario volver a realizar nuevas experimentaciones. Al final, se seleccionará el modelo que más se adecúe para llevarlo a producción.

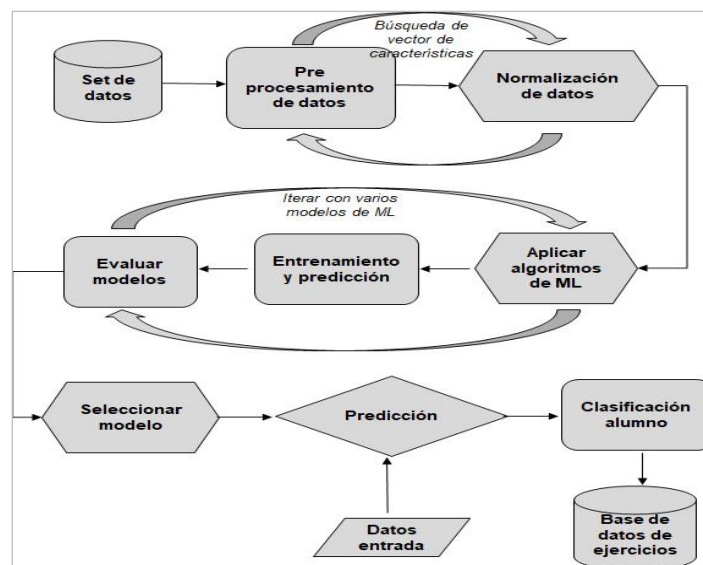


Figura 3.13: Esquema general utilizado para la selección de un algoritmo de machine learning

Capítulo 4

Análisis de datos e implementación del algoritmo de clasificación

En este capítulo, se explicarán el proceso para la preparación del set de datos, posteriormente se explicarán las diferentes experimentaciones realizadas con los algoritmos de aprendizaje automático que más se ajusten a los datos, además, se analizarán los resultados obtenidos para seleccionar un algoritmo a implementar. Finalmente, conforme se van realizando las pruebas necesarias, se comienza la estructuración del sistema web con los ejercicios necesarios para cada alumno. A continuación, se presentan y se profundizan las características antes mencionadas.

4.1 Preparación del set de datos

4.1.1 Recopilación de la información

En el capítulo anterior se habló de los principales requerimientos para implementar un sistema web que funcione con un algoritmo de machine learning. Para poder trabajar el entrenamiento del algoritmo, su proceso involucra una preparación de los datos sobre los cuales el algoritmo entrenará. En este proyecto, se busca identificar la categoría más adecuada para que un alumno pueda trabajar en el desarrollo de su proceso de alfabetización a través de una serie de ejercicios de reforzamiento por lo que es de suma importancia que los datos que se vayan a utilizar, estén lo más relacionado posible a las problemáticas actuales de alfabetización.

El colegio EDIA en la ciudad de Hermosillo, en colaboración con el sector III de la Secretaría de Educación y Cultura del Estado de Sonora proporcionaron un set de datos con un aproximado de 5300 registros que están basados en el formato general de alfabetización que se utilizaron durante el primer trimestre del ciclo escolar 2019-1, el cual, en su contenido se encontraba información general sobre los cuatro factores de recopilación de información: discapacidad, dificultades severas, trastornos y aptitudes sobresalientes, cada uno con determinadas características a identificar por parte del docente hacia el alumno.

Es importante especificar que la información entregada para este proyecto cumplió con los requerimientos de protección de datos e información, donde los datos recibidos no involucran nombres de algún niño o maestro, siendo solamente información de carácter estadístico.

En la figura 4.1 se puede observar el formato sobre el cual se trabajó la recolección de los datos por parte del docente hacia cada uno de los alumnos dentro de su aula de grupo. Como se observa, existen cuatro factores que agrupan distintas características del niño, al final, el docente asigna la categoría a cada alumno al término de cada bimestre. Las categorías sobre las que trabaja este formato son: pre silábico, silábico, silábico alfabético y alfabético.

		ESCUELA PRIMARIA: _____ C.C.T. _____ FECHA: _____																									
		DOCENTE: _____ GRUPO: _____ GRADO: _____																									
NO.	NOMBRE DEL ALUMNO	Discapacidad					Dificultades severas			Trastornos		Aptitudes Sobresalientes				AGOS./SEPT.				NOVIEMBRE							
		Intelectual	Motriz	Auditiva	Visual	Mental	Múltiple	Conducta	Comunicación	Aprendizaje	Autismo	TDA e H	Intelectual	Creativa	Sociomotriz	Artística	Psicomotriz	PS	S	SA	A	PS	S	SA	A		
1																											
2																											
3																											
4																											
5																											
6																											
7																											
8																											
9																											
10																											

Figura 4.1 Formato de alfabetización implementado por la SEC en Sonora

Uno de los principales problemas presentados durante el desarrollo del proyecto, fue el no disponer desde un inicio con un set de datos, ya que en la mayoría de los casos, las instituciones no cuentan con un registro y mucho menos una base de datos que les permita conocer las necesidades de aprendizaje de los alumnos, contando solo con el conocimiento tácito de los maestros, es decir, el conocimiento que han ido acumulando a través de la experiencia que le ha dado los años de servicio y les ha permitido sacar adelante a la mayoría de sus alumnos ante problemáticas distintas de aprendizaje.

De lo anterior, fue necesario elaborar una encuesta para cada uno de los maestros del colegio EDIA y algunas instituciones educativas que pudieran apoyar, logrando generarla con 15 preguntas. Desafortunadamente, hubo poca respuesta de los maestros, además, se observó que las respuestas de la encuesta no iban a ser suficiente información para poder entrenar un algoritmo, puesto que era necesario obtener de alguna manera el desempeño individual del alumno y no un desempeño grupal, por lo que la idea de la encuesta, paso a ser desechada. La figura 4.2 muestra parte de la encuesta subida al servidor de Google, cuyo objetivo era recabar información para integrar un set de datos.

The image shows a screenshot of a Google Forms survey. At the top, there is a decorative header with a bookshelf illustration. Below the header, the title "Proyecto EDIA" is displayed. The main body of the form contains a paragraph explaining the purpose of the survey: "La finalidad del formulario es recabar datos que puedan ser de utilidad para la implementación de un sistema de recomendación de ejercicios de lecto-escritura y conciencia fonológica a niños de primero a tercero de primaria, por lo cual es de gran importancia, que los datos que proporcione sean verídicos. Colegio EDIA agradece su colaboración para poder completar este trabajo de recopilación de información." Below this paragraph is a note: "Nota: El uso de los datos es con fines de investigación." The first question is labeled "1 - Grado actual que cursan sus alumnos *". Below the question, there is a text input field labeled "Texto de respuesta breve". On the right side of the form, there are several icons for sharing and printing.

Figura 4.2 Imagen de la encuesta para recopilación de información

Al no poder utilizar la encuesta para generar un set de datos, se optó por utilizar el formato de alfabetización entregado por el sector III, con alrededor de 5300 registros de alumnos, los cuales, mostraban en su mayoría, falta de información, esto debido a que un maestro puede identificar problemas de aprendizaje, pero no se encuentra capacitado para diagnosticar o dar veracidad sobre su propio criterio, por lo que el formato de alfabetización, con alrededor de 5300 registros, fue la opción para construir un set de datos de entrenamiento.

4.1.2 Selección del vector de características

Un vector de características, va a permitir estructurar un set de datos donde solamente se incluyan aquellos atributos que el algoritmo utilizará en su proceso de aprendizaje, por lo que la selección de estos, debe estar basada en atributos que ayuden a identificar la clase a la que pertenece el registro completo. En el caso de este proyecto, el vector se va a construir con las siguientes características: matrícula, grupo, sexo, discapacidad intelectual, discapacidad motriz, discapacidad auditiva, discapacidad visual, discapacidad mental, discapacidad múltiple, dificultad de conducta, dificultad de comunicación, dificultad de aprendizaje, trastorno de espectro autista, trastorno de déficit de atención e hiperactividad, aptitud intelectual, aptitud creativa, aptitud socio afectiva, aptitud artística y actitud psicomotriz.

Las categorías asignadas por el docente dentro del formato de alfabetización son pre silábico, para cuando el alumno aun no posee la correspondencia grafema-fonema, silábico cuando el alumno es capaz de detectar al menos un sonido en la sílaba, silábico alfabético cuando el alumno ya es capaz de detectar sonidos de sílabas y representa algunas de ellas y finalmente, alfabético cuando el alumno es capaz de detectar todos los sonidos y representa adecuadamente su letra.

El proceso anterior de categorización, se suele dar durante el primer año de educación básica, pero cuando existen problemas de aprendizaje o algún déficit en el niño, suele prolongarse hasta otros años escolares, arrastrando el problema durante su proceso de desarrollo, por lo que como se especificó en los capítulos anteriores, el algoritmo busca trabajar con alumnos durante sus primeros tres años de educación básica.

4.1.3 Proceso de normalización de datos

Una vez completado nuestro vector de características, se realizó un proceso de normalización de datos, es decir, buscar que los valores coincidan en su tipo de datos, en caso de ser solo numéricos o solo palabras, en este caso, se optó por tomar una base binaria para los datos, donde cero indica que no hay valor o no presenta alguna característica y en el caso de uno, cuando el alumno presenta alguna característica.

En el caso de las categorías, se enumeraron del uno al cuatro para los casos pre silábicos, silábico, silábico alfabético y alfabético respectivamente. La figura 4.3 muestra el set de datos normalizado y listo para someterse a entrenamiento.

do	matricula	sexo	disc_int	disc_motriz	disc_aud	disc_vis	disc_ment	disc_mult	dif_cond	difc_com	dif_aprend	trast_aut	tda-tdah	apt_int	apt_creat	apt_socioaf	apt_art	apt_psmot	clase	
1 m1		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m2		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m3		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m4		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m5		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m6		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m7		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m8		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m9		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m10		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m11		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m12		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m13		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m14		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m15		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m16		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m17		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m18		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m19		1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m20		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1
1 m21		0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	1

Figura 4.3 Set de datos normalizado para entrenamiento de algoritmo

4.2 Diseño y desarrollo del algoritmo

4.2.1 Selección y entrenamiento de algoritmos de clasificación

La parte de diseño y selección del algoritmo, busca identificar qué algoritmo se va a adaptar mejor al set de datos, para lo anterior, se han investigado los tres tipos de algoritmos de machine learning que se utilizan actualmente, donde se ha logrado identificar los algoritmos de clasificación como los más adecuados a utilizar para las experimentaciones eligiendo los siguientes cinco: KNN, SVM, Árboles de decisión, Bosques aleatorios y Naive Bayes.

A pesar de que existen más algoritmos, estos cinco seleccionados suelen ser fáciles de implementar y trabajar con poca cantidad de datos, como es en el caso del set de datos disponible.

Cabe mencionar que los cinco algoritmos suelen trabajar como algoritmos de clasificación y regresión, en este caso, se utilizarán como clasificadores.

Es importante que, durante el proceso de entrenamiento del algoritmo, el set de datos se divida donde una parte se utilice para entrenamiento y el resto para pruebas de predicciones, donde los porcentajes de división suelen estar entre valores de 20-80 o 30-70 según la cantidad de datos que se disponga y también del software utilizado para el entrenamiento. Aunque los porcentajes de división de datos nunca suelen ser los mismos para nuevos modelos, es importante probar diferentes valores para comprobar que el entrenamiento no incurra en sobreajustes (overfitting), lo que provoca que nuestro modelo haga malas predicciones, al igual que el subajuste (underfitting), que de igual forma, suele provocar malas predicciones que igual que el anterior, provocará malas predicciones, ambos suelen darse, cuando se asignan valores muy equitativos entre entrenamiento y prueba. Una forma más fácil de comprenderla, se muestra en la figura 4.4.

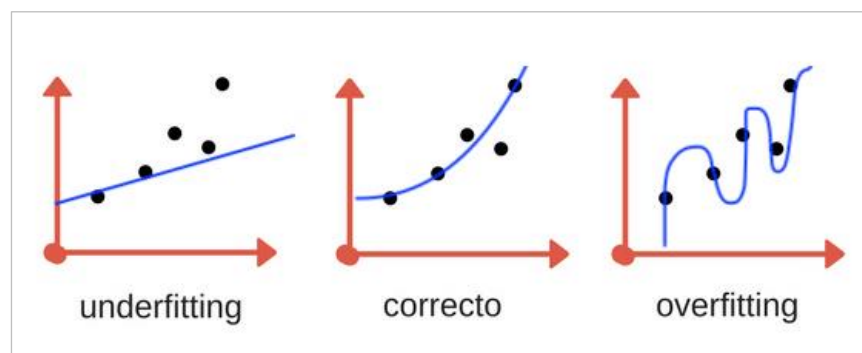


Figura 4.4 Diferencias entre underfitting y overfitting en el proceso de entrenamiento de un algoritmo de machine learning

Una vez que se ha realizado el proceso de entrenamiento y predicción se procede a evaluar los modelos y seleccionar aquel algoritmo que más se ajuste a las predicciones, por lo cual es necesario conocer el porcentaje de predicción, y en el caso de modelos de clasificación, se debe de analizar la

matriz de confusión de cada uno de ellos, esta matriz permite identificar si el modelo que se desea implementar es funcional en su predicción, ya que si se mide el modelo solamente por los aciertos que tuvo, se puede estar teniendo una falsa sensación de que el modelo funciona, es decir, se estaría sustentando la funcionalidad del modelo basándose en los falsos positivos generados.

En la figura 4.5 se puede observar el esquema de una matriz de confusión, donde cada columna de la matriz está representada por el número de predicciones de cada clase, mientras que las filas corresponden a las instancias de la clase real, su principal beneficio es facilitar si el algoritmo está confundiendo las diferentes clases o resultados de la clasificación.

		Predicción	
		Positivos	Negativos
Observación	Positivos	Verdaderos Positivos (VP)	Falsos Negativos (FN)
	Negativos	Falsos Positivos (FP)	Verdaderos Negativos (VN)

Figura 4.5: Representación de la matriz de confusión

4.2.2 Primera experimentación de los algoritmos

En esta etapa, se llevó a cabo la primera experimentación utilizando los cinco algoritmos mencionados en el capítulo anterior, además, se utilizó el set de datos de alfabetización el cual se encontraba con 5300 registros y un vector de características de 19 atributos (ya mencionados en el capítulo anterior). Se utilizó el lenguaje de programación Python, y el entorno de desarrollo Jupyter ya que el uso está basado en celdas, lo que permite la ejecución línea por línea, siendo más práctico para el trabajo con algoritmos de machine learning.

Al tener normalizado el set de datos, se realizó una primera experimentación, en la figura 4.6 se muestra parte del set de datos ya dividido para su entrenamiento el cual contiene el 80% de los datos que son utilizados.

x_train									y_train	
	disc_int	disc_motriz	disc_aud	...	apt_psmot	sex_f	sex_m		clase	
3214	0	0	0	...	0	1	0	3214	1	
4003	0	0	0	...	0	1	1	4003	4	
5011	0	0	0	...	1	0	0	5011	4	
949	0	0	0	...	0	0	0	949	2	
3961	0	0	0	...	0	1	1	3961	1	
...
4426	0	0	0	...	1	1	1	4426	3	
466	0	0	0	...	0	0	0	466	0	
3092	0	0	0	...	0	1	0	3092	1	
3772	0	0	0	...	0	0	0	3772	4	
860	0	0	0	...	0	0	0	860	2	

[4113 rows x 18 columns] [4113 rows x 1 columns]

Figura 4.6: Set de datos de entrenamiento

4.2.3 Resultados de primera experimentación de los algoritmos

Al tener el set de datos listo, se realizó la llamada a cada uno de los algoritmos estableciendo las configuraciones básicas y realizando el entrenamiento con el set de datos de la imagen anterior. Una vez realizado el entrenamiento se obtuvo la precisión de cada uno de los algoritmos, obteniendo los siguientes resultados:

Tabla 4.1: Resultados de primera experimentación

Algoritmo	Precisión (%)
KNN	36.44
SVM	49.36
Naive Bayes	13.41
Decision Trees	49.07
Random Forest	48.97

Como se puede apreciar en la tabla 4.1, se obtuvo una precisión por debajo del 50% para los cinco algoritmos y al analizar los resultados se observó que los algoritmos no se encontraban aprendiendo lo suficiente como para realizar una predicción alta, esto, debido a la gran cantidad de valores en cero en las distintas características, lo que provocó que el algoritmo no pudiera mapear un determinado valor a su posible categoría, encontrando grandes cantidades de datos con los mismos valores pero distintas clases. Además, en el capítulo anterior, hicimos la observación que parte de la información contenida en el set de datos corresponde al conocimiento tácito del maestro, ya que es este, quien llena cada uno de los formatos de alfabetización.

4.2.4 Segunda experimentación de los algoritmos

Al obtener resultados tan bajos en la primera experimentación, se procedió a realizar una reestructuración en el set de datos, donde se seleccionaron menos aquellas características que estuviesen más enfocadas a los problemas de aprendizaje, descartando las aptitudes sobresalientes, ya que no aportaban en la selección del algoritmo y al totalizar los valores de cada una de las características, se descartaron las discapacidades ya que sus valores oscilaban de 5 a 100 de un total de 5300 registros, por lo cual, en lugar de aportar al algoritmo, solo iba a provocar una baja precisión.

Para terminar con la reestructuración del set de datos, se tomaron en cuenta solamente aquellos registros que al menos contenían un valor en 1 para algún tipo de dificultad quedando el set de datos con el siguiente vector de características (matricula, grado, sexo, dificultad de conducta, dificultad de comunicación, dificultad de aprendizaje, trastorno autista y trastorno de déficit de atención) con su respectiva clasificación, con lo cual se lograron recopilar 630 registros.

4.2.5 Resultados de segunda experimentación de los algoritmos

En esta segunda parte del entrenamiento, se siguió manteniendo la configuración básica para cada uno de los algoritmos. Al entrenar de nuevo los algoritmos se obtuvieron los siguientes resultados con respecto a su precisión:

Tabla 4.2: Resultados de segunda experimentación

Algoritmo	Precisión (%)
KNN	58.73
SVM	58.73
Naive Bayes	14.28
Decision Trees	58.73
Random Forest	58.73

En la tabla 4.2 se observan los nuevos resultados, hubo un aumento aproximado el 20% pero manteniéndose apenas arriba del 50%, es decir, la predicción que los algoritmos están realizando es básicamente la probabilidad de lanzar una moneda al aire, por lo que aún no se considera listo el algoritmo a pesar de que en esta experimentación solo se utilizaron 630 datos con respecto a los 5300 registros iniciales, pero con la certeza que estos nuevos registros no contienen información en ceros. Pese a lo anterior, se considera necesaria una tercera experimentación.

4.2.6 Tercera experimentación de los algoritmos

Una vez más, al obtener resultados bajos, se optó por una tercera experimentación donde la reestructuración del set de datos fuera algo similar a la anterior, manteniendo las mismas características en el vector, pero tomando en cuenta algunas sugerencias del personal académico del colegio EDIA, basándose en los siguientes criterios para filtrar el set de datos. Al contener el set de

datos información sobre el primer bimestre del ciclo escolar, tomando en cuenta que los alumnos de primer año vienen de preescolar, los de segundo de un primer año y tercero de un segundo año cursado, se tomaron los siguientes puntos para la reestructuración:

- Se considerarán los registros de alumnos de primer año que pertenezcan a las categorías pre silábico y silábico.
- En segundo año, se incluirán las categorías silábico, silábico alfabético y alfabético, que corresponden a la mayoría de alumnos, pues consideran que son casos bastante especiales aquellos alumnos que pasan a segundo año sin saber leer ni identificar silabas.
- Para tercer año, se considerarán las categorías silábico alfabético y alfabético ya que los alumnos que presentan algún problema de aprendizaje en esta etapa están más enfocados a los últimos niveles de alfabetización.

Al terminar de aplicar los filtros recomendados por el personal del colegio EDIA, se logró estructurar un set de datos con 4722 registros.

4.2.7 Resultados de tercera experimentación de los algoritmos

En la tercera experimentación se decidió quitar el algoritmo Naive Bayes debido a que fue el único que dio resultados muy bajos, por lo cual ya no era necesario contemplarlo para esta implementación. Al finalizar el entrenamiento se obtuvieron los siguientes resultados:

Tabla 4.3: Resultados de tercera experimentación

Algoritmo	Precisión (%)
KNN	73.80
SVM	83.47
Decision Trees	82.82
Random Forest	83.04

Para esta experimentación, se puede ver en la tabla 4.3 un aumento considerable logrando obtener una precisión superior al 70% para los cuatro algoritmos, lo que, a estas instancias, permite poder utilizar el set de datos de la tercera experimentación a pesar de que la información entregada al inicio, tuvo que llevar una serie de procesos de estructuración y normalización para poder obtener datos que permitieran al algoritmo aprender, no obstante, puede ser posible realizar una cuarta experimentación solo para verificar que es posible consolidar la precisión de los algoritmos arriba de un 80%.

4.2.8 Cuarta experimentación de los algoritmos

Se ha procedido a realizar una última experimentación, en busca de aumentar un poco más la precisión de los modelos, y en vista de que es posible realizar un nuevo filtro dentro del set de datos que permita al algoritmo trabajar mejor en la predicción de las categorías 2 y 3. Tomando en cuenta lo anterior, se filtró el set de datos con las siguientes condiciones:

- Se considerarán los registros de alumnos de primer año que pertenezcan a las categorías pre silábico y silábico, en esta última, todo aquel que no tenga alguna dificultad identificada para que el algoritmo también funcione con alumnos regulares.

- En segundo año, se incluirán las categorías silábico, silábico alfabético y alfabético, siendo la categoría silábica solo para aquellos alumnos que presenten al menos una dificultad identificada por parte del docente, el resto quedará integrado como el set de datos anterior.
- Para tercer año, se considerarán las categorías silábico alfabético y alfabético, respecto la categoría silábico alfabético, solo se considerarán aquellos alumnos que tengan al menos alguna dificultad identificada por el docente.

Al terminar de aplicar los filtros recomendados se ha logrado estructurar un nuevo set de datos que contiene 4237 registros.

4.2.9 Resultados de cuarta experimentación de los algoritmos

En esta cuarta experimentación, se tomaron en cuenta los mismos cuatro algoritmos que la experimentación anterior, al finalizar el entrenamiento, se obtuvieron los siguientes resultados:

Tabla 4.4: Resultados de cuarta experimentación

Algoritmo	Precisión (%)
KNN	88.67
SVM	89.08
Decision Trees	89.15
Random Forest	89.15

En la tabla 4.4 se observan los nuevos resultados, se logró un aumento mayor con respecto a la experimentación anterior, logrando resultados por encima del 85% para los cuatro algoritmos utilizados, por lo que se puede seleccionar cualquiera de los cuatro algoritmos para llevarlo a la implementación en el sistema web.

Uno de los motivos por los cuales se optó por elegir un algoritmo dentro la cuarta experimentación esta basada en la forma en que se fueron trabajando los datos, aunque se fueron realizando filtros para ir ajustando los registros a la realidad, es importante mencionar que dentro de los algoritmos de clasificación se sabe de antemano que se puede esperar del entrenamiento, de ahí el hecho de ir realizando experimentaciones para mejorar cada vez la precisión y el aprendizaje de los algoritmos.

La tabla 4.5 muestra los resultados obtenidos de cada una de las experimentaciones que no se contemplaron debido a su bajo porcentaje de precisión, a pesar de eso, se estuvieron utilizando set de datos combinados

Tabla 4.5: Experimentaciones de los algoritmos de machine learning con diferentes sets de datos

Vector de características	Registros	KNN	SVM	AD	BA
3 dificultades, sexo, grupo	630	59.52	57.93	57.93	58.73
3 dificultades	630	51.85	53.96	53.96	52.38
3 dificultades, sexo	630	51.58	53.96	52.38	53.17
3 dificultades, grupo	630	57.93	61.11	63.49	61.90
3 dificultades, autismo	630	57.58	53.96	52.38	53.17
3 dificultades, autismo, sexo, grupo	630	59.52	60.31	57.93	57.93
3 dificultades, autismo, sexo	630	51.58	53.96	52.38	53.17
3 dificultades, autismo, grupo	630	57.93	57.93	63.49	63.49
3 dificultades, autismo, tdah	630	50.79	53.96	52.38	50.79
3 dificultades, autismo, tdah, sexo, grupo	630	58.73	58.73	58.73	59.52
3 dificultades, autismo, tdah, sexo	630	53.17	53.96	53.96	53.17
3 dificultades, autismo, tdah, grupo	630	59.52	58.73	62.69	62.69
3 dificultades, sexo, grupo	4722	57.07	53.86	53.86	53.86
3 dificultades	4722	80.01	80.95	81.37	81.37
3 dificultades, sexo	4722	54.07	53.86	53.96	54.07
3 dificultades, grupo	4722	81.03	80.95	80.95	80.95
3 dificultades, autismo	4722	54.07	53.86	53.86	53.86
3 dificultades, autismo, sexo, grupo	4722	80.58	80.95	80.53	80.53
3 dificultades, autismo, sexo	4722	41.37	53.86	53.96	53.96
3 dificultades, autismo, grupo	4722	80.69	80.95	80.95	80.95
3 dificultades, autismo, tdah	4722	28.25	53.86	53.75	53.86
3 dificultades, autismo, tdah, sexo, grupo	4722	73.80	83.47	82.82	83.26
3 dificultades, autismo, tdah, sexo	4722	41.37	53.86	54.39	54.28
3 dificultades, autismo, tdah, grupo	4722	81.16	80.95	80.74	80.74

Como es posible observar la tabla anterior, se puede ver que se utilizan como base las características de dificultades de aprendizaje, dificultades de conducta y dificultades de comunicación,

el resto de los atributos que son sexo, grupo, trastorno TDAH y autismo, se fueron combinando buscando encontrar un mejor porcentaje de precisión sobre cada entrenamiento realizado.

Además, fue durante la cuarta experimentación que se logró obtener un porcentaje de precisión superior al 85% utilizando un set de datos similar al de la tercera experimentación, excepto, en la cuarta experimentación, donde se aplicaron nuevos filtros para lograr ajustar un poco más los datos y evitar que el algoritmo trabajará con una cantidad mayor de falsos positivos.

Con lo anterior, y basándose en la última experimentación se tomó la decisión de seleccionar el algoritmo de árboles de decisión por el resto de los otros tres algoritmos principalmente tomando en cuenta su porcentaje de precisión pese a variar solo en decimales con los otros. A la par, al ser un conjunto de datos cuyos atributos pueden ser dependientes entre sí, el árbol de decisión permitió encontrar un patrón entre dichos atributos el cual se puede visualizar en el aumento de la precisión a la par con la profundidad.

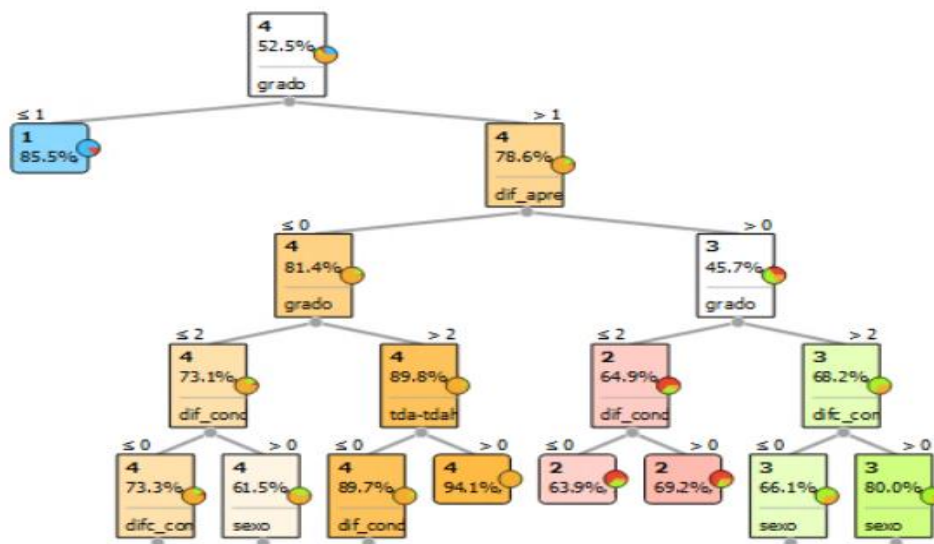


Figura 4.7: Visualización del árbol de decisión con el conjunto de datos final usando Orange

4.4 Migración del algoritmo de machine learning

Una vez que el algoritmo ya se encuentra listo para producción, se realizaron algunas pruebas desde el entorno de trabajo de Jupyter Notebook, es necesario importarlo y generar un archivo que pueda añadirse como librería en el sistema web, para lo cual, será necesaria realizar el empaquetado del archivo utilizando una librería de Python llamada Joblib, que permite exportar modelos para la futura realización de predicciones, utilizando el siguiente fragmento de código, se exportan ambos modelos por separado en un archivo con extensión .pkl:

```
In [22]: # Libreria para exportar el modelo
         from sklearn.externals import joblib
         joblib.dump(clf_dt, 'modelo_dt.pkl')

Out[22]: ['modelo_dt.pkl']
```

Figura 4.8: Código para exportar el modelo entrenado de árboles de decisión

Como se observa en la figura 4.8, el resultado del proceso es un archivo que posteriormente se utiliza como librería para realizar predicciones.

4.5 Creación de la base de datos

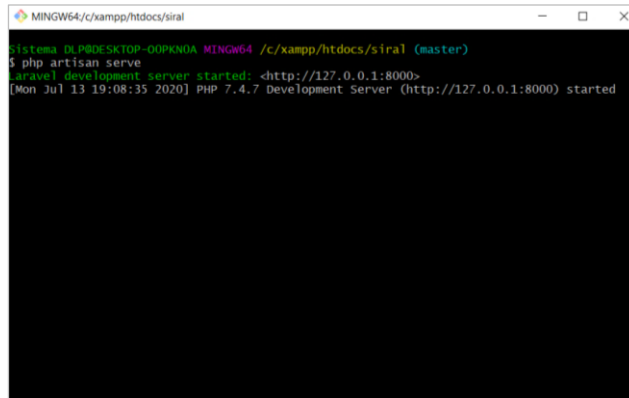
Para poder almacenar los datos tanto de entrenamiento del algoritmo, como los nuevos registros e información que alimentará al sistema, es necesario disponer de una base de datos que ayude en la gestión de los datos, por lo cual se implementará la base de datos con el servidor MySQL, la cual ayudará a persistir los datos en el sistema web.

4.6 Configuración del entorno de trabajo

Para poder disponer de un servidor web, se optó por utilizar un entorno creado en PHP configurando las librerías necesarias a través del gesto de paquetes denominado Composer, el cual permite instalar el Framework Laravel dentro del servidor local XAMPP configurado previamente para trabajar con PHP y MySQL. Para la implementación del algoritmo, se instaló un entorno de trabajo en Python y se realizó una pequeña API que permita predecir las categorías de cada uno de los alumnos creados dentro del sistema web y al consumir el servicio, se pueda recibir la relación de categoría para cada alumno, al final, será el sistema web quien se encargue de poner a disposición del alumno los diferentes ejercicios.

4.7 Instalación y configuración del Framework Laravel

Para poder comenzar a trabajar con el framework Laravel, es necesario tener instalado un servidor web, en este caso se estará trabajando con el servidor XAMPP el cual ya proporciona los servicios de Apache y MySQL. Una vez funcionando el servidor, es necesario crear un proyecto en Laravel para lo cual se necesita tener configurado el gestor de paquetes de PHP denominado Composer, y ejecutar el comando `composer create-project laravel/laravel nombre 5.8.*`. Al término, se puede levantar el servidor con la instrucción `php artisan serve`.



```
MINGW64/c/xampp/htdocs/siral
Sistema D:\PRDESKTOP-OOPKNOA MINGW64 /c/xampp/htdocs/siral (master)
$ php artisan serve
[laravel] Development server started: <http://127.0.0.1:8000>
[Mon Jul 13 19:08:35 2020] PHP 7.4.7 Development Server (http://127.0.0.1:8000) started
```

Figura 4.9: Instrucción para activar el servidor de Laravel

Es importante configurar el archivo llamado `.env` que se encuentra en la carpeta raíz del proyecto, el cual permite configurar la conexión a la base de datos, el token de seguridad para comenzar a trabajar con el framework Laravel, además de configuraciones básicas para servidor de correo electrónico, entre otras.

4.8 Programación de usuarios y CRUD

Como se pudo observar en los diagramas y figuras del capítulo anterior, se ha propuesto para el funcionamiento del sistema web el determinar tres diferentes roles de usuario, los cuales permitirán el intercambio de información y la correcta administración de datos dentro del sistema, evitando que algún usuario no acceda a funciones que corresponda a otro tipo de usuario, por lo cual, se establecieron tres roles de usuario explicados a continuación:

- Alumno: al iniciar sesión, tendrá solo acceso a su carpeta de trabajo con los diferentes ejercicios a desarrollar, pudiéndolos repetir para reforzar mas según indicaciones de su maestro.

- Maestro: un usuario maestro, tiene acceso a los CRUD (crear, leer, actualizar y eliminar) del alumno, grupos y el consumo de la API del algoritmo para evaluar a sus alumnos, además, tiene la capacidad de realizar reportes de aprovechamiento para cada uno de sus alumnos registrados.
- Administrador: tiene acceso a las herramientas del maestro, puesto que también puede desempeñar ese rol, además de coordinar los ciclos escolares y tener acceso al CRUD de maestros.

Una vez establecidos los diferentes roles de usuario ya es posible iniciar sesión y tener a disposición las opciones mencionadas anteriormente, como se puede observar en la figura 4.10 la cual muestra la pantalla de inicio del sistema una vez que se ha procedido al inicio de sesión satisfactoriamente.

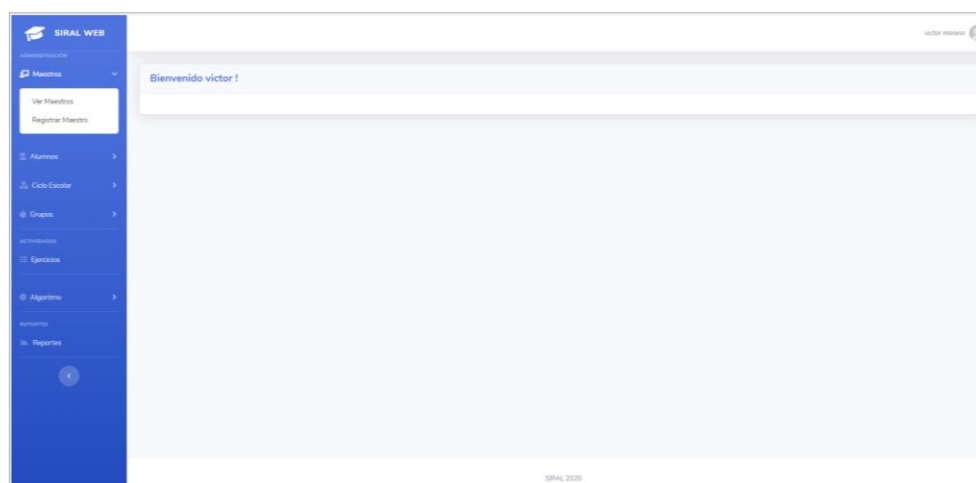


Figura 4.10: Pantalla de inicio de sesión satisfactoria

Una de las principales ventajas del manejo de roles en el sistema es la capacidad que tiene un maestro de poder administrar cómodamente sus alumnos, grupos, permitiéndole crear grupos de una

manera rápida al registrar alumnos mediante un formulario o registrar un grupo completo mediante un archivo .csv con los datos de los alumnos, en ambos casos, el sistema le asignará a cada alumno registrado un acceso que solo el maestro podrá ver y modificar en caso de que el alumno olvide o pierda su contraseña. Las dificultades registradas permitirán que el alumno pueda ser evaluado por el algoritmo de machine learning y este le pueda asignar una determinada categoría.

La figura 4.11 muestra la tabla de alumnos con tres acciones a realizar: ver las dificultades en caso de que se hayan registrado, modificar los datos del alumno, que incluyen sus datos personales y finalmente un botón para eliminar al alumno, borrando permanentemente la información almacenada en el sistema.

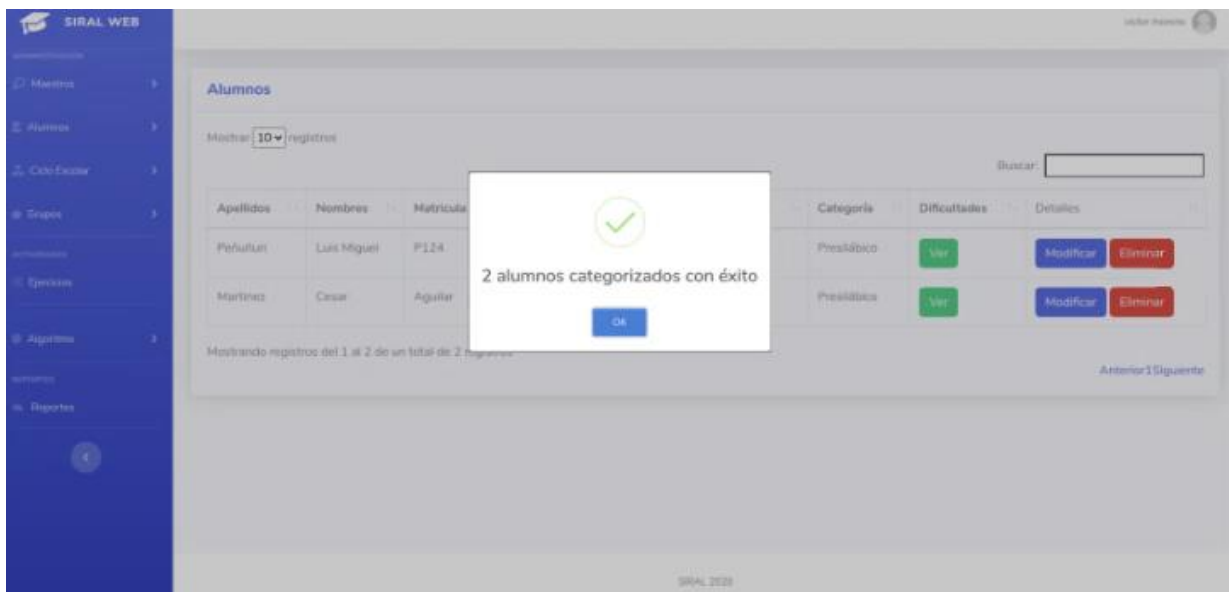


Figura 4.11: Vista de los alumnos evaluados por el algoritmo de machine learning

Capítulo 5

Análisis de resultados

En este capítulo se muestran los resultados obtenidos de las experimentaciones hechas para la selección del algoritmo y los análisis realizados para identificar la precisión ante nuevos casos clasificados por el algoritmo, además de la explicación de la matriz de confusión generada.

5.1 Resultados de las experimentaciones

Como se observó en el capítulo anterior, se tuvieron que realizar cuatro experimentaciones para determinar que se logró obtener un algoritmo con un porcentaje de precisión considerablemente bueno, pero en el proceso de cada experimentación se fueron ajustando valores y eliminando registros, todo esto debido a la presencia de datos considerados como falsos negativos, es decir, información que realmente no estaba aportando nada al set de datos lo que en las primeras experimentaciones provocó que el algoritmo no aprendiera lo suficiente arrojando resultados de predicción muy bajos. Además, el formato de alfabetización utilizado para estructurar el set de datos, solicita al docente información que en muchas ocasiones no se conoce del alumno, como lo son sus dificultades de aprendizaje, ya que incluso los tutores no suelen estar familiarizados con estos conceptos, a lo anterior, se integra el criterio de llenado empleado por el docente, provocando que al final se tenga un gran cúmulo de datos que posiblemente al someterse a algún proceso de análisis no suelen coincidir con los resultados esperados, situación que como se mencionó, compartieron los algoritmos durante las primeras dos experimentaciones.

En las últimas dos experimentaciones, se tuvo la intervención del personal del Colegio EDIA, ya que al conocer sobre las dificultades y diferentes trastornos de aprendizaje lograron estructurar un set de datos con mejor más clara, desechando registros que, por su contenido, indicaban falta de legitimidad de la situación del alumno y evitar que en estas experimentaciones existieran la menor cantidad de falsos negativos.

5.2 Análisis de la matriz de confusión

En el capítulo anterior, se llevaron a cabo cuatro experimentaciones obteniendo de la última un máximo de 89% de precisión para los algoritmos de árboles de decisión, bosques aleatorios y máquinas de soporte y aunque el algoritmo de vecinos cercanos obtuvo un 88% de precisión, se optó por analizar los de mayor porcentaje, y la selección del algoritmo para el sistema web la determinará la matriz de confusión.

Para construir la matriz de confusión se tomaron los resultados de la cuarta implementación, donde el porcentaje subió hasta un 89%, si bien se vio que los cuatro algoritmos están ofreciendo buenos resultados, es necesario aplicar otros criterios para determinar el algoritmo que va a utilizar el sistema web y determinar qué tan efectivo u óptimo puede llegar a ser el algoritmo con respecto al problema, en este caso se consideró la matriz de confusión, que como se explicó anteriormente, es una herramienta que nos permite visualizar el desempeño de un algoritmo de aprendizaje supervisado, como es el caso de los tres algoritmos con mayor precisión, donde cada columna de la matriz representa el número de predicciones para cada clase, mientras que las filas representan a los valores de la clase real. El principal beneficio de la matriz de confusión es que facilita ver si el algoritmo se encuentra confundiendo los resultados de sus predicciones.

En la figura 5.1, se pueden observar las tres matrices generadas desde el entorno de trabajo para cada uno de los algoritmos que salieron con el porcentaje de mayor precisión, como se puede observar en la imagen, la diagonal formada en la matriz indica aquellas predicciones que tuvo correcto el algoritmo y es posible ver que para la categoría 1 y 4, tuvo mayor precisión, mientras que en la categoría 2 y 3, no tuvo tanto éxito al predecir, incluso comparte grandes valores para las predicciones entre la categoría 4 de la predicción y la categoría 3 de la clase real, ocasionando en el algoritmo un poco de confusión al momento de predecir las categorías 2 y 3.

MATRIZ DE CONFUSIÓN											
Matriz SVM				Matriz Árboles Decisión				Matriz Bosques Aleatorios			
[[292	0	[[292	0	[[292	0
]]	0	0]]	0	0]]	0	0
[[49	15	[[49	16	[[49	15
]]	0	2]]	0	1]]	0	2
[[0	2	[[0	4	[[0	4
]]	7	34]]	6	33]]	5	34
[[0	2	[[0	2	[[0	2
]]	4	441]]	3	442]]	5	440

Figura 5.1: Matriz de confusión para algoritmos de cuarta experimentación

Finalmente, al observar los resultados que ofrece la matriz de confusión, es posible deducir que el algoritmo que se puede implementar en el sistema web son los árboles de decisión, a pesar de compartir el mismo porcentaje con los bosques aleatorios, muestra unas mínimas diferencias al predecir en la categoría 4 con respecto a la diagonal de los bosques aleatorios, incluyendo las clasificaciones que da hacia otra categoría, suele encontrarse con mejores precisiones que los otros dos algoritmos.

Capítulo 6

Conclusiones y trabajo a futuro

En este capítulo se presentan las conclusiones sobre el trabajo realizado y los resultados obtenidos en la implementación del algoritmo y usabilidad del sistema web. Se anexa una propuesta para trabajo a futuro, ofreciendo algunos puntos a considerar para estructurar un set de datos que permita recopilar información más específica para poder ofrecer ejercicios más específicos a cada uno de los alumnos que utilizan el sistema web.

6.1 Conclusiones

Actualmente, el uso de la tecnología para apoyo en las áreas educativas está creciendo rápidamente, es posible encontrar aplicaciones que ofrezcan apoyo para reforzar distintas áreas como las matemáticas, inglés, español, etc. Cada día las aplicaciones emplean más el uso de inteligencia artificial para determinar qué actividades deben ser consideradas para un alumno en particular, desafortunadamente en México, se está lejos de tener herramientas como las ya mencionadas en la educación básica. Si bien es cierto, las escuelas privadas pueden ofrecer estas aplicaciones de apoyo, en el caso de las escuelas públicas aun no es posible debido a la existencia de otras necesidades con mayor importancia, por lo que el desarrollo y estudio de este proyecto llevó a determinar que el sistema pueda ser utilizado desde cualquier dispositivo donde se tenga acceso a internet, en caso de no contar con el mismo, las escuelas en su gran mayoría cuentan con un aula de medios que les permite tener acceso a sistemas de ayuda ante problemas de aprendizaje.

Aunque como se mencionó en los primeros capítulos, el objetivo del sistema web no es sustituir al docente, sino, ser una herramienta de ayuda para aquellos alumnos con posibles dificultades de aprendizaje y que a través de la constante practica de ejercicios puedan mejorar sus niveles de alfabetización, en esta primera etapa del sistema, el enfoque fue en los primeros tres años de educación básica, buscando reforzar las categorías de alfabetización que se emplearon en el desarrollo del sistema, pero también es necesario el disponer de nuevas herramientas para que el alumno en su proceso de comprensión lectora, Adquiera las habilidades necesarias para no quedarse rezagado en la etapa inicial de educación.

Una de las conclusiones más importantes obtenidas en esta investigación, es la necesidad de que las instituciones públicas tengan la información suficiente que no comprometan la integridad del alumno, pero que sí permitan identificar las necesidades del mismo, y es a través de la información donde es posible emplear inteligencia artificial para encontrar los puntos a reforzar durante el proceso de aprendizaje de un alumno, de ahí la importancia de que se haga uso de la tecnología para generar información que pueda ser útil en la mejora de la educación en México.

6.2 Trabajos a futuro

El principal objetivo de seguir trabajando en este proyecto, es poder implementar en las escuelas públicas un modelo de aprendizaje a través de un sistema web que permita apoyar a los alumnos en su formación académica, ofreciéndole ejercicios no solo de lectoescritura sino también de lógica matemática según las necesidades del alumno, de ahí la importancia de crear conciencia en la generación de nuevos datos del alumno, como lo son las calificaciones, tareas, porcentajes de

aprovechamiento, etc., que den pie a la estructuración de un set de datos con información más apegada a la realidad del alumno en el aula de clases.

Referencias

APA. (2014). *Manual diagnóstico y estadístico de los trastornos mentales : DSM-5*. Editorial Médica Panamericana.

Artigas Pallarés, J. (2002). Problemas asociados a la dislexia. *Revista de Neurología*, 34(S1), 7.

<https://doi.org/10.33588/rn.34s1.2002063>

Castedo, L., Dávila, L., González, R., Hernando, M., López, S., Quesada, P., Rodríguez-Losada, D., Segundo, P. S., & Santos, Y. C. (n.d.). *ARQUITECTURA CLIENTE-SERVIDOR PARA UN LABORATORIO REMOTO*.

Castles, A., & Coltheart, M. (1993). Varieties of developmental dyslexia. *Cognition*, 47(2), 149–180.

<http://www.ncbi.nlm.nih.gov/pubmed/8324999>

Coltheart, M., Rastle, K., Perry, C., Langdon, R., & Ziegler, J. (2001). DRC: a dual route cascaded model of visual word recognition and reading aloud. *Psychological Review*, 108(1), 204–256.

<http://www.ncbi.nlm.nih.gov/pubmed/11212628>

Coltheart, Max, & Leahy, J. (1996). Assessment of Lexical and Nonlexical Reading Abilities in Children: Some Normative Data. *Australian Journal of Psychology*, 48(3), 136–140. <https://doi.org/10.1080/00049539608259520>

Critchley, M. (2015). Developmental dyslexia. In *The Cambridge Handbook of Communication Disorders* (1st Editio). William Heinemann Medical Books. <https://doi.org/10.1017/CBO9781139108683.008>

Cuetos, F. (2006). Psicología de la escritura. Diagnóstico y tratamiento de los trastornos de escritura. In *Educación al día*. Ed. Escuela Española. <https://www.casadellibro.com/libro-psicologia-de-la-escritura-diagnostico-y-tratamiento-de-los-tras-tornos-de-escritura/9788433105127/108311>

Davies, R., Cuetos, F., & Glez-Seijas, R. M. (2007). Reading development and dyslexia in a transparent orthography: A survey of Spanish children. *Annals of Dyslexia*, 57(2), 179–198. <https://doi.org/10.1007/s11881-007-0010-1>

Deuel, R. K. (1995). Developmental Dysgraphia and Motor Skills Disorders. *Journal of Child Neurology*, 10(1_suppl), S6–S8. <https://doi.org/10.1177/08830738950100S103>

DSM-IV-TR. (2002). *MSD-IV-TR*.

Escribano, B., & Enrile, M. (2017). *La dislexia causa el 40% del abandono escolar*. https://sevilla.abc.es/sevilla/sevi-dislexia-causa-40-por-ciento-abandono-escolar-201706052348_noticia.html

Farnham-Diggory, S. (2004). *DIFICULTADES DE APRENDIZAJE* (Ediciones).

Fawcett, A. J., Nicolson, R. I., Fernández-Pino, I., Corral-Gregorio, S., & Santamaría-Fernández, P. (2013). *DST-J: test para la detección de la dislexia en niños. Adaptación española* (4ta edición). TEA Ediciones.

Galaburda, A. M., & Cestnick, L. (2003). Dislexia del desarrollo. *Revista de Neurología*, 36(S1), 3. <https://doi.org/10.33588/rn.36S1.2003068>

Garfias, Á. (2012). Front-end y back-end. *Revista Software Gurú*, 36, 56–57.

Géron, A. (2017). *Hands On Machine Learning with Scikit-Learn & TensorFlow*. O'Really Media.

Giordano, L. (1978). *Discalculia Escolar: dificultades en el aprendizaje de las Matemáticas*.

Gunnulfson, M. (2013). *Scalable and efficient web application architectures: Thin-clients and SQL vs thick-clients and NoSQL* [University of Oslo]. <https://www.duo.uio.no/handle/10852/37423>

Harrington, P. (2012). Machine Learning in Action. In *Machine Learning* (Vol. 37, Issue 3). <https://doi.org/10.1007/s10994-011-5249-4>

Haugeland, J. (1988). *Inteligencia artificial*. Siglo Veintiuno Editores.

Hoién. (n.d.). *Definición y Características de La Dislexia | Dislexia | Lectura (proceso)*. Retrieved January 30, 2020, from <https://es.scribd.com/document/291214992/Definicion-y-Caracteristicas-de-La-Dislexia>

¿Puede el mecanismo de detección de errores beneficiarse del entrenamiento de la memoria de trabajo? Una comparación entre los disléxicos y los sujetos de control un estudio de ERP, (2009).

Hudson, D., Alcina, S., & Narcea, S. A. de E. (2017). *Dificultades específicas de aprendizaje y otros trastornos : guía básica para docentes*. NARCEA.

- Ibañez, A. (2018). *The Deep Learning Hype - Think Big*. <https://business.blogthinkbig.com/the-deep-learning-hype/>
- IDA. (2002). *Definition of Dyslexia – International Dyslexia Association*. <https://dyslexiaida.org/definition-of-dyslexia/>
- Jiménez, J. E. (2012). Dislexia en español : prevalencia e indicadores cognitivos, culturales, familiares y biológicos. In *Dislexia en español: prevalencia e indicadores cognitivos, culturales, familiares y biológicos, 2012, ISBN 978-84-368-2649-4, págs. 25-44*. Pirámide. <https://dialnet.unirioja.es/servlet/articulo?codigo=3897653>
- Kosc, L. (1970). Psychology and psychopathology of mathematical abilities. In *Studia Psychologica 1* (pp. 159–162).
- Méndez, F. (2018). *UNAM diseña app que detecta dislexia en niños | UNAM Global*.
<http://www.unamglobal.unam.mx/?p=33140>
- Morgan, W. P. (1896). A case of congenital word blindness. *British Medical Journal*, 2(1871), 1378.
<https://doi.org/10.1136/bmj.2.1871.1378>
- NJCLD. (1994). *Collective Perspectives on Issues Affecting Learning Disabilities. Position Papers and Statements*. Pro-Ed.
<https://acces.bibl.ulaval.ca/login?url=https://search.ebscohost.com/login.aspx?direct=true&db=eric&AN=ED385079&lang=fr&site=ehost-live>
- OMS. (2001). OMS | Informe sobre la salud en el mundo 2001 - Salud mental: nuevos conocimientos, nuevas esperanzas. *WHO*.
- Portellano, J. (2001). *La disgrafía : concepto, diagnóstico y tratamiento de los trastornos de escritura: Vol. 2{u00AA}*. Ciencias de la Educación Preescolar y Especial. https://www.worldcat.org/title/disgrafia-concepto-diagnostico-y-tratamiento-de-los-trastornos-de-escritura/oclc/758014064&referer=brief_results
- Raschka, S. (2017). *Python Machine Learning*. Packt Publishing.
- Rello, L. (2018). *Superar la dislexia : una experiencia personal a través de la investigación*. Ediciones Paidós.
http://almena.uva.es/search~S1*sp?/Ydislexia&SORT=D/Ydislexia&SORT=D&SUBKEY=dislexia/1%2C136%2C136%2CB/frameset&FF=Ydislexia&SORT=D&1%2C1%2C

- Richards, R. G. (1998). *The writing dilemma : understanding dysgraphia*. Richards Educational Therapy Center, Inc.
- Ronquillo, G. (2016). *Disgrafía: 11 síntomas del trastorno de expresión escrita*. Instituto de Neurociencias, Junta de Beneficiencia de Guayaquil. <https://www.institutoneurociencias.med.ec/component/k2/item/15037-disgrafia-sintomas-trastorno-expresion-escrita>
- Senn, J. (1995). *Análisis y diseño de sistemas de información*. McGraw-Hill.
- Share, D. L., & Stanovich, K. (1995). Cognitive Processes in Early Reading Development: Accommodating Individual Differences into a Model of Acquisition. *Issues in Education: Contributions from Educational Psychology, 1*, 1–57.
- Sierra, B. (2006). *Aprendizaje automático*. Pearson Educacion S.A.
- Small, S. L., Flores, D. K., & Noll, D. C. (1998). Different neural circuits subserve reading before and after therapy for acquired dyslexia. *Brain and Language, 62*(2), 298–308. <https://doi.org/10.1006/brln.1998.1951>
- Snow, C., Burns, M., & Griffin P. (2015). Preventing Reading Difficulties in Young Children. In *Preventing Reading Difficulties in Young Children*. National Academies Press. <https://doi.org/10.17226/6023>
- Snowling, M. J. (2000). *Dyslexia*. Blackwell Publishers.
- Vega, E. C., & Arévalo, R. G. (2011). Metodologías activas en la disgrafía. *Repositorio de La Universidad Estatal de Milagro*. <http://repositorio.unemi.edu.ec/xmlui/handle/123456789/1665?show=full>
- Winston, P. (1992). *Artificial Intelligence* (3rd.). Pearson. <https://www.amazon.com/Artificial-Intelligence-Patrick-Henry-Winston/dp/B009NGC8T8>
- Zorzi, M., Houghton, G., & Butterworth, B. (1998). Two Routes or One in Reading Aloud? A Connectionist Dual-Process Model. *Journal of Experimental Psychology: Human Perception and Performance, 24*(4), 1131–1161. <https://doi.org/10.1037/0096-1523.24.4.1131>