

# **PROTOTIPO DE TUTOR INTELIGENTE PARA EL APRENDIZAJE DE LA PROGRAMACIÓN DE COMPUTADORES**

**LUISA FERNANDA QUIROGA SAAVEDRA**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERIA DE SISTEMAS  
MODALIDAD TRABAJO DE INVESTIGACIÓN  
BOGOTÁ  
2016**

**PROTOTIPO DE UN DE TUTOR INTELIGENTE PARA EL APRENDIZAJE DE  
LA PROGRAMACIÓN DE COMPUTADORES**

**LUISA FERNANDA QUIROGA SAAVEDRA**

**Trabajo de grado para optar al título de Ingeniero de Sistemas**

**Director**

**MARIO MARTINEZ**

**M. Sc. Ingeniería de Sistemas y Computación**

**Codirector**

**ROGER GUZMÁN**

**M. Sc.(c) Ingeniería de Sistemas y Computación**

**UNIVERSIDAD CATÓLICA DE COLOMBIA  
FACULTAD DE INGENIERÍA  
PROGRAMA DE INGENIERIA DE SISTEMAS  
MODALIDAD TRABAJO DE INVESTIGACIÓN  
BOGOTÁ  
2016**



## Atribución-Compartir Igual 2.5 Colombia (CC BY-SA 2.5)

La presente obra está bajo una licencia:  
**Atribución-Compartir Igual 2.5 Colombia (CC BY-SA 2.5)**

Para leer el texto completo de la licencia, visita:  
<http://creativecommons.org/licenses/by-sa/2.5/co/>

### Usted es libre de:

- Compartir - copiar, distribuir, ejecutar y comunicar públicamente la obra
- hacer obras derivadas
- hacer un uso comercial de esta obra



### Bajo las condiciones siguientes:



**Atribución** — Debe reconocer los créditos de la obra de la manera especificada por el autor o el licenciante (pero no de una manera que sugiera que tiene su apoyo o que apoyan el uso que hace de su obra).



**Compartir bajo la Misma Licencia** — Si altera o transforma esta obra, o genera una obra derivada, sólo puede distribuir la obra generada bajo una licencia idéntica a ésta.

Nota de Aceptación

---

---

---

---

---

---

**Firma del presidente del jurado**

---

**Firma del jurado**

---

**Firma del Jurado**

**Bogotá, 24, noviembre, 2016**

## DEDICATORIA

A Dios que me ha traído en un largo camino, hasta este punto.

A mis padres. Miguel y Betsy, gracias, su apoyo, aun en las más fuertes tormentas, me dio la fuerza para seguir.

A Edwin. Amor, tú me enseñaste a creer en mi misma, fue tratando de verme con los ojos que tú me ves que encontré el que y el cómo. Tu visión me construyo.

A mi increíble hija. Laura, eres mi orgullo y mi inspiración, por ti creo en el mañana.

A María Paula. Mary, el amor que conocí al quererte no es común. Le agradezco a Dios por la oportunidad milagrosa de estar en tu vida. Laura y tú son mi esperanza y mi fuerza para querer construir un mundo mejor, un mundo para ustedes.

Tía Lucy. Como quisiera que estuviera aquí para ver este día. Mi admiración a usted es un espejo que me ha hecho más grande y más fuerte.

## **AGRADECIMIENTOS**

Son muchas las personas que debería nombrar a continuación; profesionales valiosos que en el trabajo y en la universidad han abonado mi camino como ingeniera. A todos muchas gracias por permitirme aprender y crecer académica y profesionalmente.

De forma especial quiero mencionar al profesor Mario Martínez, su paciencia, conocimiento y sabiduría fueron la llave que puso orden a un mar de ideas. Al profesor Roger Guzmán quien ha dedicado de su tiempo a guiarme por el universo de Machine Learning. Al profesor Holman Bolivar, gracias por el apoyo y por creer en mí.

A quienes de una u otra forma me prestaron su asesoría, resolviendo mis dudas y me indicaron el camino a seguir; al profesor Carlos Pulido, a los profesores de la Facultad de Psicología, Carlos Pardo y Olga Lucía Espinosa.

Al Ingeniero Edwin Hurtado, por su apoyo a nivel técnico y metodológico que me permitió superar muchos baches.

A todos, muchas gracias.

## CONTENIDO

	pág.
Nota de Aceptación	4
DEDICATORIA	5
AGRADECIMIENTOS	6
LISTA DE FIGURAS	11
LISTA DE TABLAS	13
LISTA DE ANEXOS	14
GLOSARIO	15
INTRODUCCIÓN	16
1 GENERALIDADES	18
1.1 ANTECEDENTES	18
1.2 PLANTEAMIENTO DEL PROBLEMA	18
1.2.1 Descripción del problema.	18
1.2.2 Formulación del problema.	19
1.3 OBJETIVOS	19
1.4 Objetivo general	19
1.5 Objetivos específicos	19
2 CARACTERIZACIÓN DE LA ENSEÑANZA DE LA PROGRAMACIÓN	21
2.1 COMO SE ENSEÑA	21
2.2 COMO SE APRENDE	21
2.3 QUE SE APRENDE	22
2.4 CONSTRUCTIVISMO Y METACOGNICIÓN	22
2.5 COMPETENCIA LECTORA	23
2.6 OTRAS COMPETENCIAS	23
2.6.1 Comprensión del idioma Inglés	24
2.6.2 Resolución de problemas	24
2.6.3 Concentración	24

2.6.4	Memoria	24
2.7	ANÁLISIS DE CONTEXTO	24
3	TUTORES INTELIGENTES	26
3.1	DEFINICIÓN, VENTAJAS Y LIMITACIONES	26
3.2	ARQUITECTURA	26
3.2.1	Módulo de dominio de conocimiento	27
3.2.2	Módulo del estudiante	27
3.2.3	Módulo de tutor	28
4	MODELOS MATEMÁTICOS PARA EL DISEÑO DE TUTORES INTELIGENTES	29
4.1	TEORÍA BAYESIANA	29
4.2	REDES BAYESIANAS	29
4.3	OTRAS TEORIAS	30
4.3.1	Lógica difusa	30
4.3.2	Redes neuronales	31
4.3.3	Razonamiento basado en casos	33
4.3.4	Estereotipos	34
4.3.5	Data mining	34
4.3.6	Modelo basado en reglas	35
5	ALGORITMO PARA EL RECONOCIMIENTO DE PATRONES.	36
5.1	Reconocimiento de patrones	36
5.2	Algoritmos de clasificación	36
5.2.1	K Nearest neighbors	36
5.2.2	Random forest	37

5.2.3	Support vector machines	37
5.3	Medición de desempeño	38
5.3.1	Matriz de confusión	38
5.3.2	Precisión y Recall	38
5.3.3	F1-Score	39
6	DISEÑO DEL MODELO DE DATOS PARA LAS CARACTERISTICAS	40
6.1	Evaluación de la competencia lectora	40
6.1.1	Clasificaciones de una tarea	41
6.1.2	La calificación	42
6.2	MEDICIÓN DE CONTEXTO	47
6.2.1	Edad	47
6.2.2	Estrato económico	47
6.2.3	Composición familiar	47
6.2.4	Bachillerato en colegio público o privado	48
6.2.5	Condición laboral del estudiante	48
6.2.6	Responsabilidades familiares	48
6.3	MODELO RELACIONAL PARA DATOS DEL ESTUDIANTE	48
6.3.1	Descripción de tablas	49
7	ANÁLISIS DE RESULTADOS	54
7.1	RECOLECCIÓN DE INFORMACIÓN	54
7.1.1	Contexto	54
7.1.2	Competencia lectora	54

7.2	DATOS RECOLECTADOS	55
7.2.1	Genero	56
7.2.2	Edad	56
7.2.3	Estrato socioeconómico	57
7.2.4	Composición familiar	57
7.2.5	Colegio de grado	58
7.2.6	Estudiantes trabajando	59
7.2.7	Estudiantes con hijos	59
7.2.8	Desempeño académico	60
7.3	Clasificación automática de estudiantes	61
7.3.1	K Nearest neighbors	61
7.3.2	Random Forest	64
7.3.3	Support Vector Classifier	66
8	DISEÑO Y CONSTRUCCIÓN DEL TUTOR INTELIGENTE	69
8.1	DISEÑO DEL MÓDULO DEL DOMINIO DEL CONOCIMIENTO	69
8.1.1	Red bayesiana para el curso de programación de computadores	71
8.2	DISEÑO DEL MÓDULO DEL ESTUDIANTE	74
9	CONCLUSIONES	77
10	RECOMENDACIONES	78
	BIBLIOGRAFÍA	79
	ANEXOS	83
	ANEXO A. MANUAL DE USUARIO “SISTEMA DE LEVANTAMIENTO DE INFORMACIÓN”	83

## LISTA DE FIGURAS

	pág.
Figura 1 Ejemplo de lógica difusa	31
Figura 2 Neurona	32
Figura 3 Combinación de neuronas	33
Figura 4 Razonamiento basado en casos	34
Figura 5 Ejemplo de planos de decisión en SVM	37
Figura 6 Ecuación de precisión	38
Figura 7 Ecuación de recall	38
Figura 8 Ecuación F1 score	39
Figura 9 Escala para la competencia lectora	43
Figura 10 Modelo relacional para datos del estudiante	49
Figura 11 Cuestionario preguntas de contexto	54
Figura 12 Ejemplo de estímulo	55
Figura 13 Ejemplo de pregunta	55
Figura 14 Estudiantes por género	56
Figura 15 Edades de estudiantes	56
Figura 16 Estrato de estudiantes	57
Figura 17 Composición Familiar Estudiantes	58
Figura 18 Colegio Estudiantes	58
Figura 19 Estudiantes que trabajan	59
Figura 20 Estudiantes con hijos	60
Figura 21 Desempeño académico	60
Figura 22 Dataset	61
Figura 23 Entrenamiento del modelo de KNeighborsClassifier	62
Figura 24 Predicción del modelo	62
Figura 25 Inicialización de la matriz de confusión	62
Figura 26 Matriz de confusión Neighbors Classifier	63
Figura 27 Reporte de clasificación KNeighbors Classifier	64
Figura 28 Entrenamiento para Random Forest	64
Figura 29 Predicción para Random Forest	65
Figura 30 Matriz de confusión para Random Forest	65
Figura 31 Reporte de clasificación para Random Forest	66
Figura 32 Entrenamiento SVC	66
Figura 33 Predicción para SVC	67
Figura 34 Matriz de confusión para SVC	67
Figura 35 Reporte de clasificación para SVC	68
Figura 36 Subgrafo para el concepto de ciclo For	70
Figura 37 Red bayesiana propuesta	73
Figura 38 Ejemplo de construcción de la red en Jayes	74



## LISTA DE TABLAS

	Pág.
Tabla 1 Niveles de desempeño de lectura	43
Tabla 2 Tabla de estudiantes	49
Tabla 3 Tabla de competencias	50
Tabla 4 Tabla de niveles	51
Tabla 5 Tabla de tareas	52
Tabla 6 Tabla de preguntas	52
Tabla 7 Tabla Nivel estudiante	53
Tabla 8 Probabilidad total para el concepto de ciclo For	70
Tabla 9 Conceptos para red bayesiana	72

## LISTA DE ANEXOS

Anexo A	pág. 83
---------	------------

## GLOSARIO

**APRENDIZAJE:** Adquisición del conocimiento de algo por medio del estudio, el ejercicio o la experiencia, en especial de los conocimientos necesarios para aprender algún arte u oficio.

**CARACTERÍSTICA:** Cualidad o circunstancia que es propia o peculiar de una persona o una cosa y por la cual se define o se distingue de otras de su misma especie.

**COGNICIÓN:** Capacidad del ser humano para conocer por medio de la percepción y los órganos del cerebro.

**COMPETENCIA:** Las competencias son los conocimientos, habilidades, y destrezas que desarrolla una persona para comprender, transformar y practicar en el mundo en el que se desenvuelve.

**CONSTRUCTIVISMO:** Teoría explicativa de los procesos de aprendizaje a partir de conocimientos ya adquiridos.

**INFERENCIA:** Acción de extraer un juicio o conclusión a partir de hechos, proposiciones o principios, sean generales o particulares.

**MACHINE LEARNING:** Se trata de crear programas capaces de generalizar comportamientos a partir de una información no estructurada suministrada en forma de ejemplo.

**METACOGNICIÓN:** Se refiere a la capacidad de las personas para reflexionar sobre sus procesos de pensamiento y la forma en que aprenden. Gracias a la metacognición, las personas pueden conocer y regular los propios procesos mentales básicos que intervienen en su cognición.

## INTRODUCCIÓN

La dinámica económica y social del mundo manifiesta una creciente demanda de personal especializado en tecnologías de la información y telecomunicaciones (TIC), en particular de desarrolladores de software. En un país como Colombia, con una población de 45 millones de personas, había un déficit de 15000 ingenieros y profesionales con este perfil para el 2014<sup>1</sup>, y se espera llegue a 100000 para el 2019<sup>2</sup>, es por esto que la formación de dichos profesionales es una prioridad.

Iniciativas como la del Ministerio TIC, quien ha designado una inversión de 50000 millones de pesos en el plan Vive Digital 2014-2018 para incentivar con becas el estudio de Ingeniería de Sistemas y afines<sup>3</sup>, pero aun así la formación de profesionales en desarrollo de software enfrenta problemas más allá de la financiación, la motivación del estudiante se ve reducida frente a las numerosas y variadas habilidades que debe reunir a la hora de programar.

El estudiante novato, que toma un primer curso de programación, tiene carencias que se relacionan con múltiples aspectos de su historia académica y que pueden ser tan variados como la capacidad de entender un texto, analizar y resolver un problema, y plantear una estrategia.

El éxito en superar estos obstáculos parece estar relacionada con los dos actores principales de este proceso, como son el estudiante y el profesor; el primero al tener la constancia suficiente para insistir en los aspectos que representan un reto y el segundo al identificar dichos aspectos y plantearle al estudiante un plan que le permita superarlo con la frustración mínima.

---

<sup>1</sup> EL ESPECTADOR. “Hay déficit de 15.000 ingenieros” para industria TIC: Mintic [en línea]. Colombia: Economía [citado >27 de octubre, 2015]. Disponible en Internet: <http://goo.gl/awS10S>

<sup>2</sup> MINISTERIO TIC COLOMBIA EN FACEBOOK. [En línea]. Colombia: Ministerio TIC Colombia [citado 27 de octubre, 2015]. Disponible en Internet: <https://goo.gl/W6Es6c>

<sup>3</sup> REVISTA SEMANA. ¿Y dónde están los ingenieros? [En línea]. Colombia: Tecnología [citado 27 de octubre, 2015]. Disponible en Internet: <http://goo.gl/rTE5qh>

Esta dinámica estudiante – profesor se ve obstruida por el tiempo y el volumen de estudiantes por clase. En este trabajo se buscan estrategias para aumentar la efectividad del tiempo de estudio y la influencia del profesor, se analizan dos perspectivas, la primera, la forma en que las cualidades y circunstancias del estudiante, es decir, sus características influyen en su asimilación del curso; luego se examina la forma en que plan de estudio de la materia debe ser abordado para asegurar que un tema no se inicie hasta que se garantice un nivel de comprensión de los temas prerrequisito.

El marco de este trabajo está limitado a los estudiantes de ingeniería de la Universidad Católica de Colombia

# 1 GENERALIDADES

## 1.1 ANTECEDENTES

Las investigaciones que se han seguido en los últimos 100 años alrededor de los procesos educativos han revolucionado la forma en que, tanto profesores como alumnos interactúan con la dinámica de la educación, es así como, por ejemplo, se pasó de tener un componente totalmente magistral donde el alumno solo era un testigo a ser un proceso donde profesor y alumno trabajan de la mano.

Aunque la investigación en la enseñanza es un campo amplio, es necesario profundizar en los factores que influyen en el aprendizaje de un campo como el de la programación de computadores para poder llegar a conclusiones que permitan solventar más efectivamente las dificultades que enfrentan los estudiantes y así aumentar el éxito del sistema educativo.

Con relación al desarrollo de tutores inteligentes este es un tema que se ha tratado desde el nacimiento de la Inteligencia Artificial y la documentación de la efectividad del uso de algoritmos como Redes Bayesianas y *Bayesian knowledge tracing* es lo suficientemente sólido como para usar esta herramienta como un punto de partida para el desarrollo de un sistema cada vez más inteligente y más empático.

## 1.2 PLANTEAMIENTO DEL PROBLEMA

### 1.2.1 Descripción del problema.

La popularidad de los sistemas computacionales ha cambiado el mundo. Son muchos los dispositivos de cómputo con los que una persona se ve involucrada todos los días, desde lavadoras y neveras inteligentes hasta celulares y videojuegos, esto hace que la industria de las TIC se haya convertido en una de las más importantes del mundo, lo que incide en la mayor demanda de desarrolladores de software, pero la realidad es que existe un déficit de profesionales en esta área, lo que ha puesto al sistema educativo frente al reto de capacitar los profesionales que demanda la sociedad.

En las facultades de ingeniería enfrentan la falta de iniciativa y la deserción en sus programas de sistemas. La aproximación tradicional a la enseñanza de desarrollo de software en las aulas de clase los profesores pueden observar como sus estudiantes se sienten desmotivados y frustrados a la hora de desarrollar, esto debido a malas bases en su educación previa, a los tamaños de las clases, dado

que el profesor no puede poner en cada uno la atención individual que se requiere e incluso la situaciones personales.

Este panorama requiere un análisis que debe comenzar por identificar las características que son determinantes en el éxito de un estudiante en un curso de programación de computadores para poder apoyar su proceso de aprendizaje en los puntos donde es más importante y así evitar la frustración y la desmotivación.

Como un primer acercamiento a la solución del problema antes mencionado se analizan las características particulares del estudiante que permitan establecer los factores de influencia en el éxito de un estudiante.

Luego el diseño y desarrollo del prototipo de un tutor inteligente para la enseñanza y aprendizaje de la programación de computadores que le permita al estudiante avanzar a su ritmo por medio del diseño de un plan de estudio que se adapte a sus necesidades, su nivel de conocimientos y su facilidad de aprendizaje.

### **1.2.2 Formulación del problema.**

- ¿Cuáles son las características que afectan el desempeño del estudiante en el curso?
- ¿Cuáles son las relaciones de causalidad que permitan predecir que estudiantes tienen factores de riesgo que afectarán su desempeño?
- ¿Cómo organizar el conocimiento de forma eficiente y dinámicamente para diseñar la mejor ruta de aprendizaje para cada estudiante?
- ¿Cómo establecer el estado de conocimiento de un estudiante en programación de computadores?

### **1.3 OBJETIVOS**

#### **1.4 Objetivo general**

Desarrollar un tutor inteligente para apoyar el proceso de aprendizaje de programación de computadores usando técnicas de Aprendizaje de Maquina.

#### **1.5 Objetivos específicos**

- Determinar los criterios de representación de características de los estudiantes en el aprendizaje de programación de computadores.

- Construir una base de datos que permita almacenar la representación de características personales y cognitivas de los estudiantes.
- Identificar la técnica basada en aprendizaje de máquina para la construcción del tutor inteligente de programación de computadores.
- Diseñar e implementar un método automático basado en Aprendizaje de Máquina para categorizar a los estudiantes según las características personales y cognitivas.

## **2 CARACTERIZACIÓN DE LA ENSEÑANZA DE LA PROGRAMACIÓN**

Se asume que al aprender a programar, el estudiante desarrolla ciertas habilidades, pero los profesores encuentran que no todos los estudiantes las desarrollan con la misma eficiencia porque tienen deficiencias en aptitudes y actitudes subyacentes.

### **2.1 COMO SE ENSEÑA**

La enseñanza de la programación de computadores en el sistema educativo actual toma el grupo de estudiantes como una entidad con una dinámica única, pasando un tema tras otro sin garantizar que cada estudiante haya alcanzado un estado de manejo del tema que le de las herramientas suficientes para abordar el siguiente, lo cual causa que progresivamente el alumno se vea superado por el que hacer de la materia y termine por darse por vencido y alejarse definitivamente de la programación.

### **2.2 COMO SE APRENDE**

Las habilidades con las que debe contar un programador son de naturaleza amplia; la programación abarca nociones tan variadas como la inteligencia emocional y la estrategia de solución de problemas. Estas son habilidades que, en su mayoría, no se pueden desarrollar en un solo semestre de práctica académica y debe ser un conjunto subyacente con que el estudiante cuente al iniciar el estudio.

Las habilidades que se pueden destacar incluyen:

- Establecer un plan detallado: Cuando un programador enfrenta la solución de un problema, el primer paso luego de la comprensión del propio problema es proponer los pasos lógicos que deberá seguir para obtener la solución.
- Conocer el lenguaje de programación: Para que el plan pueda ser implementado con éxito, el programador debe conocer las características que ofrece el lenguaje.

- Hacer “*debugging*”: A la hora de confirmar lo que el programa hace, o lo que no está haciendo, es importante poder interpretar mensajes obtenidos de salidas de error y logs y en general rastrear el flujo de la aplicación.<sup>4</sup>

### 2.3 QUE SE APRENDE

El curso de programación de computadores de la Universidad Católica de Colombia busca el desarrollo de competencias generales y específicas en el estudiante, como son:

- Capacidad de solucionar problemas.
- Capacidad de analizar y sintetizar.
- Capacidad de aprender y actualizarse.

De forma específica la asignatura busca que el estudiante sea capaz de abstraer de la realidad los elementos de un problema que impliquen la automatización de tareas para el manejo de información.

Analizar, modelar y describir los componentes necesarios para la solución de problemas con computador.

Diseñar e implementar en un lenguaje de programación algoritmos eficientes que garanticen la calidad de la información en la solución del problema.<sup>5</sup>

### 2.4 CONSTRUCTIVISMO Y METACOGNICIÓN

Entender la forma en que los individuos aprenden ha sido objetivo de la psicología pedagógica por años; en particular los conceptos del constructivismo y la metacognición se presentan como particularmente útiles para entender como el estudiante se relaciona con lo que aprende a través de lo ya aprendido.

La metacognición puede definirse simplemente como la capacidad de “pensar sobre pensar”. Es la forma en que las personas aprenden a aplicar el pensamiento

---

<sup>4</sup> BBC ACTIVE. Why you should use a Virtual Learning Environmen [en línea]. [citado 29 de octubre, 2015]. Disponible en Internet: <http://goo.gl/d8YjDP>

<sup>5</sup> UNIVERSIDAD CATÓLICA DE COLOMBIA. Diseño asignatura programación de computadores CT10047

a la forma de actuar y de aprender del entorno, para lo cual se usa una reflexión constante para reutilizar los conocimientos adquiridos.<sup>6</sup>

El constructivismo dice que las personas construyen su propio entendimiento y conocimiento del mundo, a través de experimentar las cosas y reflexionar en esas experiencias. Cuando se encuentra algo nuevo, se intenta reconciliarlo con las ideas y experiencias previas, ya sea cambiando una creencia o descartando un dato nuevo como irrelevante.<sup>7</sup>

El constructivismo ha tenido gran influencia en la enseñanza de ciencias y matemáticas, lo cual puede aprovecharse para tener una base teórica que los use para entender la forma en que se aprende a programar y rastrear las habilidades y situaciones personales que son significativas en el proceso y que pueden ser determinantes en el éxito de un estudiante.<sup>8</sup>

## **2.5 COMPETENCIA LECTORA**

Si bien es cierto que estas causas subyacentes podrían ser muchas se tomará como referencia la comprensión de lectura.

Resultados previos sugieren una correlación entre la lectura y el desempeño matemático, centrándose en los cursos de segundo a quinto, donde incluso la correlación crece entre más alto es el grado.<sup>9</sup>

La relación entre comprensión de lectura y el desempeño académico parece lógica dada la alta importancia de la información textual en nuestra sociedad.

## **2.6 OTRAS COMPETENCIAS**

El conjunto de habilidades que debe manejar un estudiante para tener un acercamiento efectivo a la programación de computadores, se encontrará mucho más amplio de lo analizado en el presente trabajo. Para obtener un mayor impacto

---

<sup>6</sup>DEFINICIÓN.DE. Definición de metacognición. <http://definicion.de/metacognicion/>

<sup>7</sup>CONCEPT TO CLASSROOM. Costructivism as a Paradigm for teaching and learning <http://www.thirteen.org/edonline/concept2class/constructivism/>

<sup>8</sup> BEN-ARI, Mordechai. Constructivism in computer science education. *Journal of Computers in Mathematics and Science Teaching*, 2001, vol. 20, no 1, p. 45-74.

<sup>9</sup> SIEMENS, George. *Connectivism: A learning theory for the digital age*. 2014.

y establecer un plan de preparación más efectivo se deberían tener en cuenta entre otras competencias, las siguientes.

### **2.6.1 Comprensión del idioma Inglés**

El volumen de documentación en el área de programación de computadores lleva a los estudiantes que no manejan un nivel adecuado a tener que conformarse con las fuentes en español.

### **2.6.2 Resolución de problemas**

Para la algoritmia, parte importante de la programación de computadores y en general para la ingeniería, la resolución de problemas abarca una serie de técnicas que busca lograr una meta partiendo de una condición presente.

### **2.6.3 Concentración**

La habilidad en focalizar la atención en un solo pensamiento o tarea, excluyendo todo lo demás de la conciencia es llamada concentración. Estudios realizados en diferentes ámbitos muestran que las personas con alto nivel de desempeño manejan esta capacidad.<sup>10</sup>

### **2.6.4 Memoria**

Dado que para poder programar se debe construir un mapa mental del problema a tratar, sus entradas y sus salidas, se observará como la memoria juega un papel en la capacidad de avanzar en el plan de estudios.

## **2.7 ANÁLISIS DE CONTEXTO**

Un análisis socioeconómico, visto como la medida de educación, ingreso y ocupación busca situar al estudiante en un contexto. Investigaciones indican que los alumnos con un bajo status socio-económico desarrollan habilidades académicas más lentamente comparados con estudiantes de status más altos.<sup>11</sup>

---

<sup>10</sup> GOLEMAN, Daniel. Focus, el motor oculto de la excelencia.

<sup>11</sup>AMERICAN PSYCLOGICAL ASSOCIATION. Education and Socieconomics Status. <http://www.apa.org/pi/ses/resources/publications/education.aspx>

Se busca establecer como este fenómeno afecta el estudio de la programación de computadores.

-

## 3 TUTORES INTELIGENTES

### 3.1 DEFINICIÓN, VENTAJAS Y LIMITACIONES

El propósito de los tutores inteligentes (Intelligent Tutoring System, ITS) no es comunicarse con el estudiante en la misma forma que lo haría un profesor humano, sino más bien comunicar el conocimiento efectivamente.

En términos generales el propósito de un tutor inteligente es establecer el conocimiento que se quiere enseñar y exponer este conocimiento al estudiante de forma progresiva y limitada a lo que este está listo a asimilar, asegurando así que el estudiante no sea subestimado, lo que causará que se aburra o sobreestimado, con lo cual el conocimiento lo abrumara.

Las técnicas tradicionales para los tutores inteligentes pueden ser clasificadas en tres grupos que son: secuenciadores de currículos, interacción para el soporte en la solución de problemas y análisis inteligente de las soluciones del estudiante.

Los secuenciadores de currículos y el análisis inteligente de la solución del estudiante son las tecnologías más viejas y mejor estudiadas en el dominio de ITS. La tecnología de interacción para el soporte en la solución de problemas es la más nueva, pero también la más inteligente, pues ayuda al estudiante en el proceso más difícil del aprendizaje.

Sin embargo para el contexto web se hacen más adecuadas las aproximaciones más antiguas, en particular los secuenciadores de currículos que pueden guiar al estudiante en medio de la gran cantidad de información disponible en línea.<sup>12</sup>

### 3.2 ARQUITECTURA

En los estudios previos del campo se encuentran numerosos módulos incluidos en el diseño de un posible tutor, dependiendo de la orientación y objetivo que tenga, pero en general cuenta con las siguientes partes o módulos principales:

---

<sup>12</sup> DEPARTMENT OF COMPUTER SCIENCE PRINCETON UNIVERSITY. What is machine learning?, theoretical machine learning [en línea], Rob Schapire. [citado 29 de octubre, 2015]. Disponible en Internet: <http://goo.gl/WHZ1jq>

### 3.2.1 Módulo de dominio de conocimiento

Contiene la información o habilidad en la que el estudiante será instruido.<sup>13</sup> Debe organizar el material de estudio, representado como artículos, notas de clase, videos o cualquier material que pueda ayudar a entender un concepto; también debe organizar los ejemplos prácticos que ilustran los conceptos y los casos de evaluación que para este caso se centran en preguntas de selección múltiple con única y múltiple respuesta y que presentan un enunciado, el cual puede incluir un concepto o un caso práctico y las respuestas, resaltando la correcta. Tanto los quizzes, como el material de estudio están organizados por conceptos, lo que en la fase del tutor permite que el material se indexe y recupere con mayor facilidad.

Es aquí donde el módulo del tutor encuentra la información necesaria para decidir qué ruta sugerir al estudiante y para esto se debe poder establecer qué sabe el estudiante, que no sabe y que está listo para aprender, lo cual incluye manejar una gran cantidad de incertidumbre, pues el estudiante podría responder correctamente una pregunta asociada a un concepto por error, adivinar, o puede responder mal aunque si maneja el concepto, incluso puede creer que entiende y de hecho no hacerlo.

### 3.2.2 Módulo del estudiante

Se encarga de establecer el nivel de conocimiento del estudiante y monitorear su progreso. Es importante que haga seguimiento de su estilo de aprendizaje y de sus preferencias pedagógicas.

Se articula con el Modelo del Dominio del Conocimiento, pues lo asiste al permitir tener en cuenta características del estudiante y no solo del conocimiento, como la probabilidad de que realmente haya aprendido el concepto. Otros factores, como su estado afectivo o el nivel de complejidad en los problemas que está listo a afrontar, no se tienen en cuenta pues se ha demostrado que incluir más variables hace el modelo demasiado complicado y tiene un impacto muy bajo al aumentar la precisión.<sup>14</sup>

---

<sup>13</sup> PILLAY, Nelishia. Developing intelligent programming tutors for novice programmers. *ACM SIGCSE Bulletin*, 2003, vol. 35, no 2, p. 78-82.

<sup>14</sup> IMAGINE BLOG. Bayesian Knowledge Tracing [en línea]. Matthew W. Johnson. Julio 2012. [citado 29 de Octubre de 2015]. Disponible en Internet <http://goo.gl/rj10sF>

### **3.2.3 Módulo de tutor**

Recibe información del módulo del estudiante y del módulo del dominio para establecer una estrategia de tutoría. Este módulo es el que regula la interacción del estudiante con el conocimiento. También es el encargado de decidir los métodos de evaluación y retroalimentar al módulo del estudiante.

La primera parte de este módulo selecciona el conocimiento que el estudiante está listo para aprender, lo que se logra poniendo los conceptos en tres categorías, conocido, listo para conocer y no listo para conocer. Un concepto se asume conocido si acumula una probabilidad de 1, si tiene una por debajo de uno y todos los conceptos padres son conocidos se marca listo para aprender, por último, si por lo menos un concepto padre no es conocido, se deja no listo para aprender.

## 4 MODELOS MATEMÁTICOS PARA EL DISEÑO DE TUTORES INTELIGENTES

### 4.1 TEORÍA BAYESIANA

El teorema de Bayes, en la teoría de la probabilidad, es una proposición planteada por el filósofo inglés Thomas Bayes (1702-1761), que expresa la probabilidad condicional de un evento aleatorio A dado B en términos de la distribución de probabilidad condicional del evento B dado A y la distribución de probabilidad marginal de sólo A.

El teorema de Bayes se expresa por la siguiente fórmula:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

En términos más generales y menos matemáticos, el teorema de Bayes es de enorme relevancia puesto que vincula la probabilidad de A dado B con la probabilidad de B dado A. Es decir, que sabiendo la probabilidad de tener un dolor de cabeza dado que se tiene gripe, se podría saber (si se tiene algún dato más), la probabilidad de tener gripe si se tiene un dolor de cabeza. Muestra este sencillo ejemplo la alta relevancia del teorema en cuestión para la ciencia en todas sus ramas, puesto que tiene vinculación íntima con la comprensión de la probabilidad de aspectos causales dados los efectos observados.

### 4.2 REDES BAYESIANAS

Una red bayesiana, red de Bayes, red de creencia, modelo bayesiano (de Bayes) o modelo probabilístico en un grafo acíclico dirigido es un modelo grafo probabilístico (un tipo de modelo estático) que representa un conjunto de variables aleatorias y sus dependencias condicionales. Por ejemplo, una red bayesiana puede representar las relaciones probabilísticas entre enfermedades y síntomas. Dados los síntomas, la red puede ser usada para computar la probabilidad de la presencia de varias enfermedades.

Formalmente, las redes bayesianas son grafos dirigidos acíclicos cuyos nodos representan variables aleatorias en el sentido de Bayes: las mismas pueden ser cantidades observables, variables latentes, parámetros desconocidos o hipótesis. Las aristas representan dependencias condicionales; los nodos que no se encuentran conectados representan variables las cuales son condicionalmente independientes de las otras. Cada nodo tiene asociado una función de probabilidad que toma como entrada un conjunto particular de valores de las variables padres del nodo y devuelve la probabilidad de la variable representada por el nodo. Por ejemplo, si por padres son variables booleanas entonces la función de probabilidad puede ser representada por una tabla de entradas, una entrada para cada una de las posibles combinaciones de los padres siendo verdadero o falso. Ideas similares pueden ser aplicadas a grafos no dirigidos, y posiblemente cíclicos; como son las llamadas redes de Markov.

Existen algoritmos eficientes que llevan a cabo la inferencia y el aprendizaje en redes bayesianas. Las redes bayesianas que modelan secuencias de variables (ej. señales del habla o secuencias de proteínas) son llamadas redes bayesianas dinámicas. Las generalizaciones de las redes bayesianas que pueden representar y resolver problemas de decisión bajo incertidumbre son llamados diagramas de influencia.

## 4.3 OTRAS TEORIAS

### 4.3.1 Lógica difusa

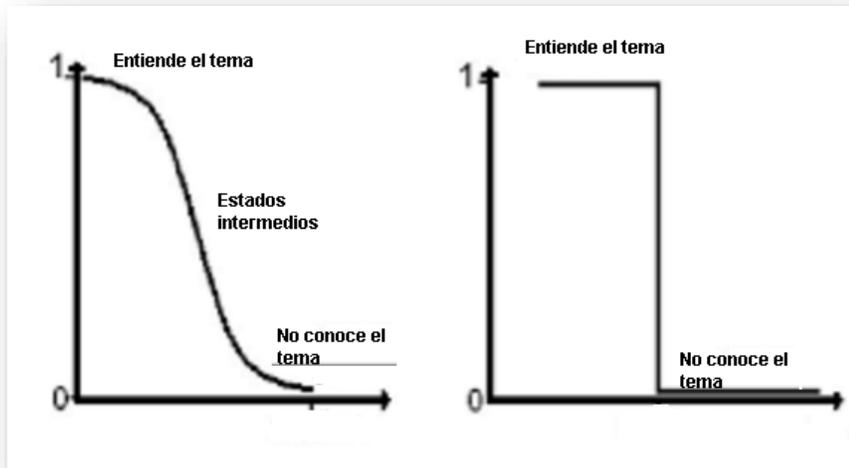
El término “lógica difusa” surge en el desarrollo de la teoría de conjuntos difusos de Lofti Zadeh en 1965. Un subconjunto difuso  $A$  de un conjunto  $X$  está caracterizado por asignar a cada elemento  $x$  de  $X$  un grado de pertenencia de  $x$  en  $A$ . Por ejemplo, si  $X$  es un grupo de estudiantes  $A$  es un subconjunto difuso de estudiantes de  $X$  que dominan un tema. Los elementos de  $X$  tendrán asignados un grado de verdad, que puede ir desde es “absolutamente falso” hasta “absolutamente cierto” o cualquier nivel de verdad intermedio<sup>15</sup> Ver Figura 1 Ejemplo de lógica difusa. Entender perfectamente el tema y no tener idea son solamente dos estados en los extremos de todas las posibles relaciones con un

---

<sup>15</sup> Stanford Encyclopedia of Philosophy. Fuzzy Logic  
<http://plato.stanford.edu/entries/logic-fuzzy/>

concepto, como son: entiende la idea básica, entiende ejemplos, puede aplicarlo, entre otros.<sup>16</sup>

Figura 1 Ejemplo de lógica difusa



Fuente: El autor

Esta técnica ha sido usada para mejorar el desempeño de un ITS dada su versatilidad para manejar información imprecisa, tal como las acciones del estudiante, que pueden ser explicadas en más de una forma posible, y para proveer una descripción del conocimiento y de las habilidades del estudiante más ajustada a la forma en que lo haría un humano.

#### 4.3.2 Redes neuronales

Una Red Neuronal Artificial (Artificial Neural Network ANN) es un paradigma de procesamiento de información inspirado en la forma como el sistema nervioso biológico procesa información. La estructura de procesamiento está compuesta por un gran número de elementos altamente interconectados, llamados neuronas, trabajando al unísono para resolver un problema específico. Estos sistemas, al igual que las personas, aprenden por ejemplos. Una red neuronal está configurada

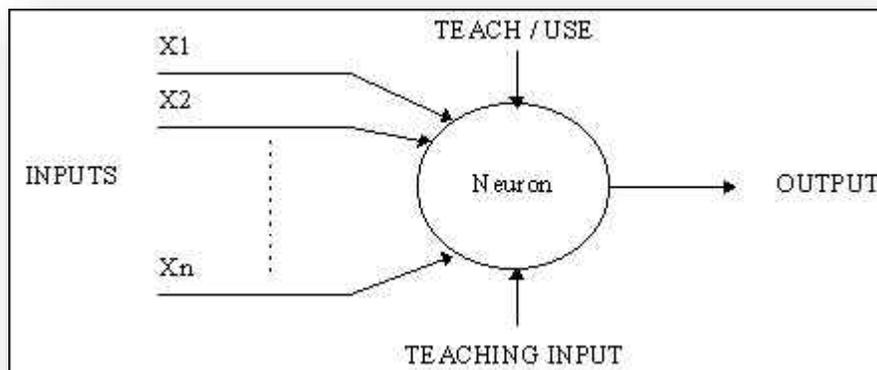
---

<sup>16</sup> STATHACOPOULOU, R.; MAGOULAS, G. D.; GRIGORIADOU, M. Neural network-based fuzzy modeling of the student in intelligent tutoring systems. En *Neural Networks, 1999. IJCNN'99. International Joint Conference on*. IEEE, 1999. p. 3517-3521

para el reconocimiento de patrones o la clasificación de datos, a través del aprendizaje de procesos. El aprendizaje en sistemas biológicos involucra la calibración entre las conexiones sinápticas existentes entre las neuronas.<sup>17</sup>

Una neurona es una unidad con muchas entradas y muchas salidas. Tiene dos modos de operación, modo de entrenamiento y modo de uso. En el modo de entrenamiento la neurona es entrenada para responder o no, a un patrón de entrada en particular. En el modo de uso, cuando un patrón de entrada enseñado es detectado se retorna la salida asociada. Si el patrón de entrada no corresponde a ningún en la lista de entradas la regla de respuesta es usada para determinar si se responde o no. Ver Figura 2 Neurona

Figura 2 Neurona



Fuente: Imperial London College  
[https://www.doc.ic.ac.uk/~nd/surprise\\_96/journal/vol4/cs11/report.html](https://www.doc.ic.ac.uk/~nd/surprise_96/journal/vol4/cs11/report.html)

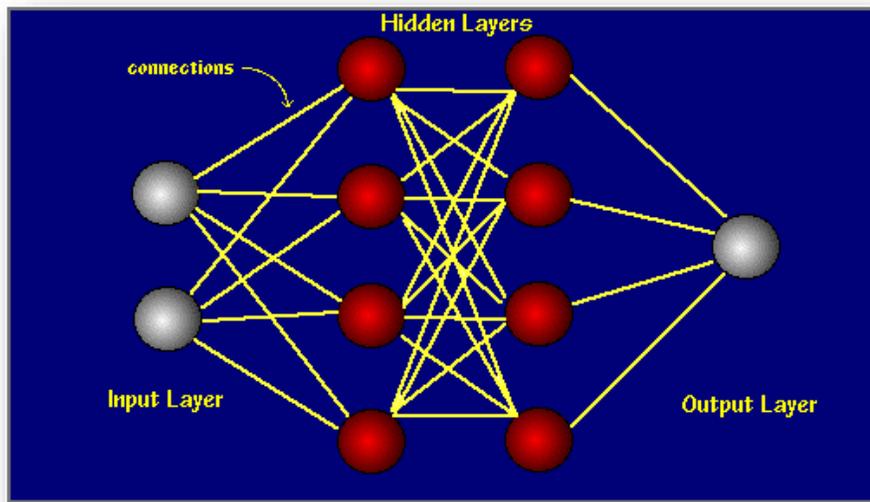
La regla de respuesta se basa en la distancia Hamming, que es el número de posiciones en las que dos símbolos son diferentes, y así la neurona responde a un patrón de entrada para el que no fue entrenada según sea la respuesta para el patrón más cercano al actual.

Las neuronas se combinan para encontrar información oculta en los datos de entrada y devolver diferentes conclusiones, Figura 3 Combinación de neuronas

---

<sup>17</sup> Imperial College London. Introduction to Neural Networks. <http://tinyurl.com/zf8alvk>

Figura 3 Combinación de neuronas



Fuente: Computer Science User Pages.  
<http://pages.cs.wisc.edu/~bolo/shipyard/neural/local.html>

Se han usado en el diseño de ITS tanto para simular el proceso cognitivo del estudiante como un control para adaptar la simulación del estudiante. Las redes neurales están entrenadas para imitar las decisiones respecto a las características del estudiante que tomaría un profesor humano

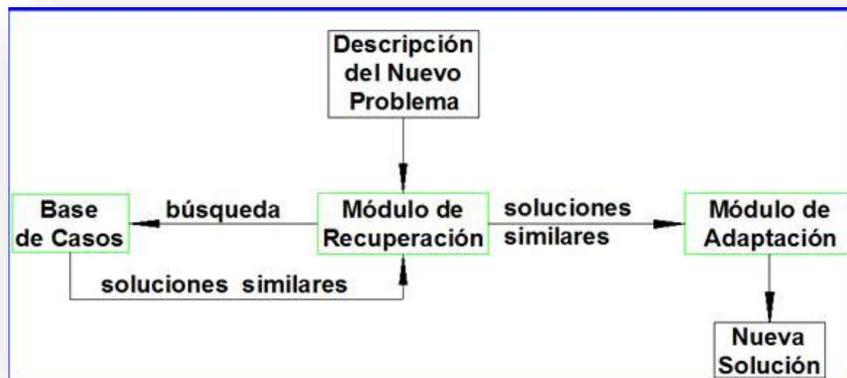
#### 4.3.3 Razonamiento basado en casos

La idea es adaptar una solución exitosa a un problema previo a la solución de un nuevo problema. Por la rapidez en su desarrollo se cuenta como una de las herramientas de inteligencia artificial más efectivas, dado que una base de datos existente puede ser usada como fuente de conocimiento. Las mesas de ayuda y los sistemas de diagnóstico son las aplicaciones más comunes.

Normalmente el razonamiento basado en casos usa algoritmos de recuperación y de correspondencia a una base de pares de problemas y soluciones basadas en casos. Los casos son frecuentemente derivados de bases de datos legadas.

Tiene la ventaja de ser adaptable a dominios donde hay muchas formas en la cual una regla puede ser aplicada, por lo tanto existen muchos caminos para llegar a una respuesta correcta. En los casos aparece un sistema típico

Figura 4 Razonamiento basado en casos



Fuente: Razonamiento basado en casos: Cesar Alberto Chagoyen Mendez <http://www.monografias.com/trabajos98/razonamiento-basado-casos-via-diseno-engranes/razonamiento-basado-casos-via-diseno-engranes.shtml>

#### 4.3.4 Estereotipos

Es un mecanismo útil para construir modelos de usuarios individuales con base en una cantidad pequeña de información. La principal idea de los estereotipos es agrupar todos los posibles usuarios de un sistema adaptativo en diferentes grupos de acuerdo con ciertas características que comparten típicamente.

Por ejemplo, alguien pregunta a una bibliotecaria por libros sobre China. ¿Qué tipo de libros debería entregarle la bibliotecaria? Depende. ¿La persona es un niño que vio un programa sobre China en televisión y quiere ver más fotos sobre China y sus lugares exóticos? ¿La persona es un estudiante haciendo un informe sobre China? ¿O tal vez es un futuro turista? ¿O un estudiante interesado en las corrientes de pensamiento Oriental? ¿Sabe chino la persona? La bibliotecaria necesita saber estas cosas antes de poder ofrecer a la persona el libro correcto, así que construye un conjunto de conclusiones sobre la persona basado en su primer vistazo al sujeto.

#### 4.3.5 Data mining

*Data mining* es el proceso de analizar datos desde diferentes perspectivas y resumirlo en información útil. Usa algoritmos matemáticos para segmentar los datos y evaluar la probabilidad de eventos futuros.

El problema se aborda construyendo modelos. Un modelo usa un algoritmo que trabaja sobre un conjunto de datos. Existen diferentes formas de *data mining*, que apuntan por ejemplo a predicción, donde, un modelo puede predecir los ingresos basado en la educación y otros factores demográficos. Las predicciones tienen una probabilidad asociada. Otra forma es el agrupamiento, que identifica las agrupaciones naturales en los datos, por ejemplo, podría identificar el segmento de la población que tiene sus ingresos en un rango específico, que tiene un historial de conducta bueno y que estrena carro en un análisis anual.

En el campo de la enseñanza involucra el desarrollo, investigación y métodos de computación aplicados a detectar patrones en la información de los estudiantes en grandes cantidades de información de educación que de otra forma serían imposibles o muy difíciles de analizar por el enorme volumen de datos.

#### **4.3.6 Modelo basado en reglas**

Un sistema basado en reglas consiste de tres componentes básicos: un conjunto de reglas (base de reglas), una base de datos (base de hechos), y un intérprete de reglas. En su diseño más simple una regla es un par ordenado de cadenas de símbolos. La regla tiene un orden total predeterminado y la base de datos es una colección de símbolos. El intérprete es diseñado para buscar cada lado de las reglas hasta que alguno concuerda con la base de datos.<sup>18</sup>

Es un conjunto de reglas que deben ser seguidas por el estudiante paso a paso durante el desarrollo del problema. Cada acción recibe una retroalimentación de acuerdo al método correcto de ejecutar la tarea. Este modelo tiene cuatro puntos principales: (1) sugiere al estudiante el siguiente paso a seguir, (2) da demostraciones, (3) evalúa el conocimiento que el aprendiz procesa en término de las habilidades que son aplicadas, y (4) infiere los objetivos del aprendiz.

---

<sup>18</sup> NWANA, Hyacinth S. Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 1990, vol. 4, no 4, p. 251-277.

## 5 ALGORITMO PARA EL RECONOCIMIENTO DE PATRONES.

### 5.1 Reconocimiento de patrones

El reconocimiento de patrones es un campo que tiene que ver con el descubrimiento de regularidades en la información a través del uso de algoritmos y valerse de estas regularidades para tomar acciones como clasificar la información en categorías.<sup>19</sup>

Con el objetivo de procesar la información, se organiza en conjuntos  $N = \{x_1, x_2 \dots x_n\}$ , donde cada  $X_n$  es un caso particular fenómeno a estudiar. Un grupo de los datos se toma como conjunto de entrenamiento y es usado para ajustar el modelo, este grupo está acompañado de un vector objetivo (*target vector*), que contiene las categorías en las que se puede clasificar el conjunto de entrenamiento y que corresponde a cada  $X_n$ .

El algoritmo puede expresarse como  $y(x)$ , donde la forma de  $y$  se obtiene de la fase de entregamiento y cuya salida corresponde con el vector objetivo. Luego de obtener el modelo, se puede usar para clasificar nueva información, al aplicar la función  $y$  a este conjunto de prueba, se obtendrá una clasificación para cada caso consecuente con el conjunto objetivo.

### 5.2 Algoritmos de clasificación

Cuando el conjunto de entrenamiento contiene vectores de entrada con su correspondiente vector objetivo se habla de una rama del reconocimiento de patrones llamada aprendizaje supervisado y en particular, cuando el objetivo del problema es asignar cada vector de entrada a una de un conjunto discreto de categorías se habla de problemas de clasificación.

#### 5.2.1 K Nearest neighbors

En la fase de entrenamiento de este método se almacenan todos los ejemplos de prueba, con su etiqueta y luego, cuando se va a hacer la predicción se calcula la distancia del elemento de prueba a cada ejemplo de entrenamiento. Se escogen

---

<sup>19</sup> BISHOP, Christopher M. Pattern recognition. *Machine Learning*, 2006, vol. 128.

los  $k$  vecinos más cercanos, con  $k$  como un entero y la etiqueta más común entre estos vecinos será la predicción para el caso de entrenamiento.<sup>20</sup>

### 5.2.2 Random forest

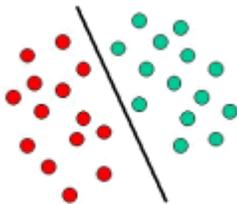
Este método se basa en árboles de decisión, en los cuales se crean árboles para predecir el valor de una variable con respecto a variables de entrada y donde cada nodo del árbol corresponde con una variable y el valor de la variable de destino esa representado por el camino desde la raíz a la hoja.

*Random forest* genera muchos árboles de decisión y para clasificar un vector de entrada lo aplica a los diferentes árboles generados. Cada árbol da una clasificación y esta se toma como un “voto” por esa etiqueta. El algoritmo escoge la etiqueta que tenga más “votos”.<sup>21</sup>

### 5.2.3 Support vector machines

Se basan en la definición de fronteras de decisión por medio de planos, estos separan conjuntos que tienen diferentes etiquetas. Como se muestra en la figura 5 los elementos que pertenecen a la clase rojo o verde, la línea que los separa define una frontera en la que del lado derecho todo es verde y al izquierdo todo es rojo, así, cualquier objeto nuevo, que llegue y se ubique en el lado izquierdo será clasificado como rojo o verde si queda del lado derecho.<sup>22</sup>

Figura 5 Ejemplo de planos de decisión en SVM



Fuente: <http://www.statsoft.com/Textbook/Support-Vector-Machines>

---

<sup>20</sup> ELKAN, Charles. Nearest neighbor classification. *elkan@cs.ucsd.edu*, 11 January, 2011, vol. 11, p. 3.

<sup>21</sup> Department of Statistics, University of California, Berkeley. Random forest

<sup>22</sup> STATISTICA, Support Vector Machines

## 5.3 Medición de desempeño

### 5.3.1 Matriz de confusión

La matriz de confusión contiene información sobre la clasificación real y la realizada por un sistema de predicción. El desempeño de tales sistemas es comúnmente evaluado con los datos en una matriz, donde las columnas representan el número de predicciones de cada clase, mientras las filas representan las clasificaciones reales.<sup>23</sup>

### 5.3.2 Precisión y Recall

Precisión es la fracción de casos obtenidos que son relevantes, mientras que Recall hace referencia al porcentaje de casos relevantes que son obtenidos.

La precisión se calcula con el número de verdaderos positivos ( $T_p$ ) dividido por la suma de verdaderos positivos y falsos positivos ( $F_p$ ) como en la Figura 6 Ecuación de precisión

Figura 6 Ecuación de precisión

$$P = \frac{T_p}{T_p + F_p}$$

Fuente: [http://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_precision\\_recall.html](http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html)

La recall es el número de verdaderos positivos ( $T_p$ ) sobre el número de verdaderos positivos más el número de falsos negativos ( $F_n$ ), como en la Figura 7 Ecuación de recall.<sup>24</sup>

Figura 7 Ecuación de recall

$$R = \frac{T_p}{T_p + F_n}$$

---

<sup>23</sup> Department de Computer Science. University of Regina. Confusion Matrix. [http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion\\_matrix/confusion\\_matrix.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html)

<sup>24</sup> Scikit learn. Precision-Recall. [http://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_precision\\_recall.html](http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html)

Fuente: [http://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_precision\\_recall.html](http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html)

### 5.3.3 F1-Score

Se interpreta como el peso promedio de la precisión y recall. Figura 8 Ecuación F1 score

Figura 8 Ecuación F1 score

$$F1 = 2 * (precision * recall) / (precision + recall)$$

Fuente: [http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1\\_score.html](http://scikit-learn.org/stable/modules/generated/sklearn.metrics.f1_score.html)

## **6 DISEÑO DEL MODELO DE DATOS PARA LAS CARACTERISTICAS**

### **6.1 Evaluación de la competencia lectora**

Tal como lo destaca la Organización de las Naciones Unidas (ONU), la competencia lectora, en conjunto con la escritura y el cálculo numérico, es pieza clave para garantizar la participación plena de un individuo en la sociedad, y se presenta tanto como un resultado de la educación, como un instrumento para continuar aprendiendo, en particular en un mundo tan cambiante como el actual, además que facilita el ejercicio de otros derechos.

Por lo anterior y debido a que el proceso de aprendizaje y resolución de problemas, en particular en entornos virtuales se hace a través de medios escritos; se plantea medir el nivel de lectura en que se encuentran los participantes del presente proyecto, con el fin de evaluar la relevancia del mismo en su proceso de aprendizaje.

Esta competencia mide la capacidad de un individuo para comprender, utilizar, y reflexionar sobre los textos escritos y comprometerse con ellos para alcanzar sus propios objetivos, desarrollar sus conocimientos y su potencial, y participar en la sociedad.

Como marco para realizar esta medición se ha seleccionado el sugerido por las pruebas PISA (Programa para la evaluación internacional de alumnos).

Las pruebas PISA surgen en 1997, como iniciativa de la Organización para la Cooperación y el Desarrollo Económico (OCDE) y su objetivo es medir el nivel de preparación que tiene un estudiante de 15 años para enfrentar su rol en la sociedad; y al ser aplicada desde al año 2000 y cada tres años, PISA también busca evaluar la evolución de los sistemas educativos de los países involucrados.

La pertinencia en seleccionar PISA como marco para la medición de esta competencia se resalta en que hace énfasis no solo en la capacidad para extraer datos de un texto, sino que se extiende a la capacidad de integrar información del mismo y compararla con conocimientos previos.

La competencia lectora se define como:

“Competencia lectora es comprender, utilizar y reflexionar sobre textos escritos para alcanzar los propios objetivos, desarrollar el conocimiento y potencial personales, y participar en la sociedad.”<sup>25</sup>

Esto abarca desde la decodificación básica hasta el conocimiento de palabras, gramática y estructura de los textos, donde no solo se busca comprender sino también utilizar lo que se lee y reflexionar en cuanto a que es un proceso interactivo.

Las pruebas PISA buscan diferentes resultados, como son:

- Indicadores contextuales, que muestran la relación de los logros de aprendizaje con variables demográficas, económicas, sociales y educativas. El cual el presente trabajo ya ha buscado abordar previamente.
- Indicadores de aprendizaje: El cual provee un perfil base de los conocimientos y los habilidades de los estudiante. Este es el resultado que se busca obtener en esta parte.
- Indicadores de tendencia: Los cuales aprovechan el carácter y buscan establecer una evolución. Este indicador, debido a la duración del presente estudio, será postergado para futuros analisis.

Dentro de las pruebas PISA el conocimiento se mide en unidades que constan de un estímulo, el cual puede ser un texto, una tabla, un gráfico, un diagrama; seguido por una serie de preguntas o tareas asociadas a un estímulo en común.

### 6.1.1 Clasificaciones de una tarea

Si bien en las pruebas PISA se encuentran diversas clasificaciones en las que puede situarse una tarea, en este caso solo se tomarán en cuenta las siguientes:

Formato del texto: Los distingue entre continuos y discontinuos. Los textos **continuos** están formados a su vez por oraciones que a su vez se organizan en párrafos. Los textos **discontinuos** están formados por listas, gráficos, diagramas, anuncios, horarios, catálogos, índices y formularios. Los textos **mixtos** están formados por textos como continuos como discontinuos

La prueba mide tres competencias lectoras principales:

---

<sup>25</sup> OCDE. Evaluación de las competencias lectoras para el siglo XXI. Pág 21

- Acceder y obtener: Obtener información, que centra al lector en fragmentos de información aislados dentro del texto
- Integrar e interpretar: Desarrollar una comprensión global y elaborar una interpretación centran al lector en las relaciones dentro del texto. Las tareas que se centran en todo el texto exigen que los lectores desarrollen una comprensión global
- Reflexionar y evaluar: reflexionar sobre el contenido de un texto y evaluar sobre la forma de un texto.

### **6.1.2 La calificación**

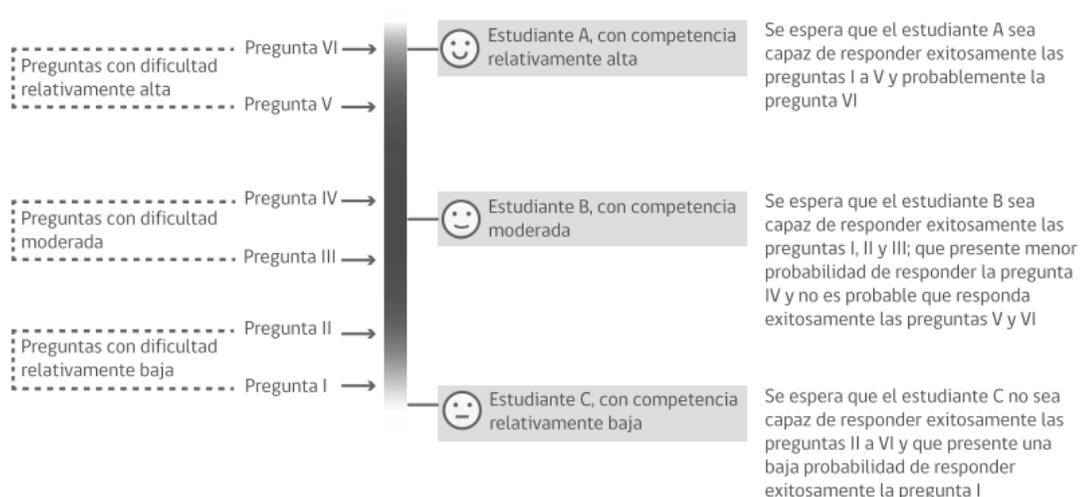
Las tareas de lectura se distribuyen a lo largo de una escala que indica, de forma progresiva, el nivel de dificultad para los estudiantes y el nivel de competencia requerido para responder a cada tarea correctamente. La escala resume tanto la competencia de una persona en función de su capacidad, como la complejidad de una pregunta en función de su dificultad. Es decir, es posible asociar la competencia de cada estudiante con un punto particular en la escala de competencia lectora, que indica su nivel estimado de competencia y es posible asociar cada pregunta con un punto particular en la misma escala, que indica su dificultad.

Esto significa que es probable que los estudiantes sean capaces de responder correctamente las preguntas que están en o bajo el nivel de dificultad asociado con su propia posición en la escala.

A pesar que la prueba original cuenta con diferentes tipos de preguntas, incluyendo preguntas abiertas; para el presente estudio sólo se usará el modelo de preguntas de selección múltiple con única respuesta. Esto implica que la calificación se da entre “no obtiene puntuación” y obtiene el puntaje completo de la pregunta. En la prueba completa se encuentran respuestas que tienen puntaje parcial.

En la Figura 9 Escala para la competencia lectora, se observa la escala que relaciona las preguntas, las tareas y los estudiantes a través del desempeño obtenido en las mismas en las mismas.

**Figura 9 Escala para la competencia lectora**



Fuente: OCDE. Evaluación de las competencias lectoras para el siglo XXI

La escala general está basada en todas las preguntas, cuenta con siete niveles, Tabla 1 Niveles de desempeño de lectura, y lo que indican es que un estudiante en un nivel 2 será capaz de responder las preguntas de nivel 2 e inferior, pero es poco probable que responda exitosamente tareas de nivel superior. Es así como un estudiante en nivel 6 será, más probablemente, capaz de responder todas las preguntas de la prueba.

**Tabla 1 Niveles de desempeño de lectura**

Nivel	Puntaje inferior	límite	Características
1a	262		Las tareas en este nivel requieren que el lector localice una sola información presentada de manera explícita y en una posición destacada en un texto breve, de sintaxis sencilla con un tema y tipo de texto conocidos, como una narrativa sencilla o una lista simple. El texto generalmente brinda apoyo al lector, mediante repetición de información, imágenes o símbolos conocidos. Hay una cantidad mínima de información que compete. En las tareas que requieren interpretación, el lector puede necesitar hacer conexiones

		simples informaciones adyacentes
1b	335	Las tareas en este nivel requieren que el lector localice una o más informaciones presentadas de manera explícita, para reconocer el tema principal o la intención del autor en un texto de tema conocido o que haga una conexión simple entre la información del texto y el conocimiento cotidiano común. Generalmente la información se destaca en el texto y hay poca o ninguna información que compite. El lector es dirigido explícitamente para que considere factores relevantes en la tarea y en el texto.
2	407	Algunas tareas en este nivel requieren que el lector localice una o más informaciones que deben ser inferidas y deben cumplir con diversas condiciones. Otras implican reconocer la idea principal de un texto, comprender relaciones o construir significados dentro de una parte limitada del texto cuando la información no destaca y el lector debe hacer inferencias de bajo nivel. Las tareas en este nivel pueden involucrar comparaciones o contrastaciones basadas en una sola característica del texto. Las tareas típicas de reflexionar en este nivel requieren que los lectores hagan una comparación o diversas conexiones entre el texto y conocimiento externo al texto, recurriendo a la experiencia y actitudes personales.
3	580	Las tareas en este nivel requieren que el lector localice diversa información que debe cumplir con múltiples condiciones y en algunos casos que reconozca la relación entre la información. Las tareas de

		<p>interpretación requieren que el lector integre distintas partes del texto para identificar la idea principal, comprender una relación o construir el significado de una palabra o una frase. Ellos deben tener en cuenta muchas características para comparar, contrastar o categorizar. A menudo, la información requerida no destaca o hay mucha información que compite; o existen otros obstáculos en el texto, como ideas contrarias a lo que se espera o formuladas como negaciones. Las tareas de reflexión en este nivel pueden requerir establecer conexiones, comparaciones y explicaciones, o requerir que el lector evalúe una característica del texto. Algunas tareas de reflexionar requieren de los lectores una comprensión exacta de textos relacionados con conocimientos que son familiares o cotidianos. Otras tareas no demandan una comprensión detallada, pero sí que el lector use conocimientos menos comunes.</p>
4	553	<p>Las tareas en este nivel que requieren obtener información demandan al lector localizar y organizar información de difícil acceso. Algunas tareas en este nivel requieren interpretar sutilezas del lenguaje de una sección del texto, teniendo en cuenta el texto como un todo. Otras tareas interpretativas requieren comprender y aplicar categorías en un contexto desconocido. Las tareas de reflexionar requieren que los lectores usen conocimiento formal o informal para elaborar hipótesis o realizar una evaluación crítica de un texto. Los lectores deben demostrar una comprensión exacta de textos largos y complejos cuyo contenido o forma puede no ser familiar.</p>

5	626	Las tareas en este nivel que requieren obtener información demandan al lector localizar y organizar diversa información de muy difícil acceso, infiriendo cuál información es relevante. Las tareas de reflexión demandan que el lector recurra a conocimientos específicos, para realizar una evaluación crítica o elaborar hipótesis. Tanto las tareas de interpretación como las de reflexión requieren una comprensión global y detallada de un texto cuyo contenido o forma no es familiar. Para todos los aspectos de lectura, las tareas en este nivel implican manejar conceptos que pueden ser contrarios a las expectativas
6	698	Las tareas en este nivel requieren que el lector haga inferencias múltiples, comparaciones y contrastaciones que sean detalladas y precisas. Requieren la demostración de una comprensión completa y detallada de uno o más textos y pueden requerir la integración de información de más de un texto. Las tareas pueden requerir que el lector trabaje con ideas que no son familiares para él, con información destacada que compite, y que genere categorías abstractas de interpretación. Las tareas de reflexión y evaluación pueden requerir que el lector, a partir de un texto complejo sobre un tema que no le es familiar, elabore hipótesis o haga una evaluación crítica, teniendo en cuenta diversos criterios o perspectivas, y aplicando una interpretación sofisticada que vaya más allá del texto. Una característica que destaca en las tareas de acceder y obtener en este nivel es la precisión de análisis y la atención a detalles no destacados en los textos.

Fuente: OCDE. Evaluación de las competencias lectoras para el siglo XXI

## **6.2 MEDICIÓN DE CONTEXTO**

Al ser estas variables las que afectan al sujeto de forma externa y sobre las que se tiene poca influencia, se usarán para situar a los individuos en grupos y ver cómo estas influyen en su proceso de aprendizaje.

### **6.2.1 Edad**

Según los niveles de conocimiento de Piaget el proceso de desarrollo de un individuo relaciona el crecimiento y el aprendizaje, por un lado habla sobre factores biológicos y por el otro de procesos de aprendizaje que se desarrollan a partir de la interacción entre el individuo y el entorno, lo que se conoce como proceso cognitivo. Jean Piaget, reconoció que la fuerte relación entre ambos procesos, biológico y cognitivo, pues, por ejemplo, es imposible dar con un niño de dos meses que haya tenido dos años para interactuar con el ambiente. El modelo de Piaget llega hasta la adolescencia, sin embargo se quiere evaluar si en la edad adulta el proceso de aprendizaje sigue evolucionando con la edad y como esto lo afecta.

### **6.2.2 Estrato económico**

Al indagar en los factores asociados al rendimiento académico se aborda con interés la condición socioeconómica pues esta ofrece un marco explicativo y en los resultados se puede apoyar la construcción de políticas institucionales más adecuadas para los estudiantes. Aunque su influencia directa en el desempeño académico no parece ser determinante<sup>26</sup>, es importante establecer cómo afecta esto al desempeño particular en programación de computadores.

### **6.2.3 Composición familiar**

La relación entre el clima familiar y el clima escolar, entendida como la actitud hacia la autoridad, la actitud empática, es un tema de estudio en los colegios. En particular se busca observar de qué forma la conformación del núcleo familiar afecta a los estudiantes, ya que esta es una variable de más fácil medición.

---

<sup>26</sup> VARGAS, Guiselle María Garbanzo. Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17 (3), 2013, p. 57-87.

#### **6.2.4 Bachillerato en colegio público o privado**

Al analizar los diferentes factores socioeconómicos que influyen en el desempeño académico se hace necesario incluir al colegio del que procede el estudiante debido a su influencia en la preparación para la educación superior. Se aprovecha este apartado para analizar como la titularidad del colegio afecta el rendimiento del alumnado. Tomando como base la información recolectada en las pruebas PISA, se ha observado que los centros privados obtienen, en general, mejores resultados que los públicos, es este resultado relevante para el curso de programación de computadores.<sup>27</sup>

#### **6.2.5 Condición laboral del estudiante**

Se busca analizar cómo afecta el desempeño del estudiante el dividir su tiempo entre las responsabilidades laborales y las académicas y cuál es la diferencia cuando no las hay, pues en la comunidad es común ver una porción significativa de estudiantes que trabajan.

#### **6.2.6 Responsabilidades familiares**

En relación con la característica de la condición laboral del estudiante también se busca observar cómo el rol que el estudiante desempeña dentro de su núcleo familiar afecta sus resultados académicos.

### **6.3 MODELO RELACIONAL PARA DATOS DEL ESTUDIANTE**

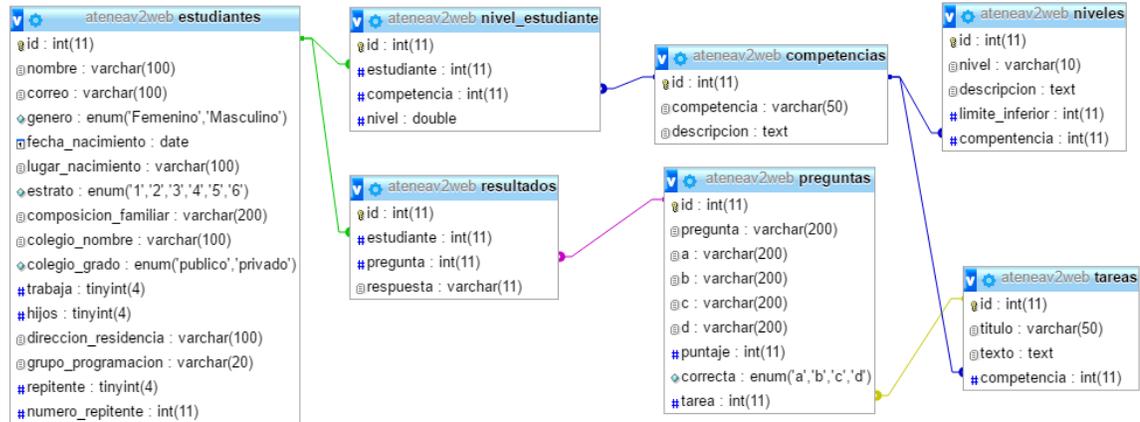
Para poder almacenar la información recolectada de forma ordenada y que favorezca el análisis de información. Para este propósito se seleccionó una base de datos relacional, usando como motor MariaDB 10.1.8.

El diagrama relacional es el propuesto en la Figura 10 Modelo relacional para datos del estudiante

---

<sup>27</sup> EL CONFIDENCIAL. ¿Colegios públicos o privados? El informe PISA analiza cuáles son mejores. <http://tinyurl.com/ptdIrlz>

**Figura 10 Modelo relacional para datos del estudiante**



Fuente: El autor

### 6.3.1 Descripción de tablas

#### 6.3.1.1 Estudiantes

En esta tabla se registra la información personal del estudiante y su información de contexto. Esta información permitirá establecer la influencia de estos factores en el desempeño académico.

**Tabla 2 Tabla de estudiantes**

Campo	Tipo	Descripción
id	Int – Llave primaria	Identificador único del alumno
Nombre	Varchar	Nombre del alumno
Correo	Varchar	Correo electrónico de la universidad
Genero	enum('Femenino', 'Masculino')	Género del estudiante

Fecha_nacimiento	date	Se usa para determinar su edad
Lugar_nacimiento	varchar	Lugar de nacimiento del estudiante
Estrato	enum('1', '2', '3', '4', '5', '6')	Estrato socioeconómico del estudiante
Composición_familiar	varchar	Descripción de la estructura familiar del estudiante
Colegio_nombre	Varchar	Colegio de grado
Trabaja	tinyint	Estado laboral del estudiante
Hijos	tinyint	Almacena si el estudiante tiene hijos
Dirección_residencia	Varchar	Lugar de residencia del estudiante
Grupo_programacion	varchar	Grupo de programación del estudiante
Repitente	tinyint	Ha repetido la materia
Nuemero_repitente	int	Cuántas veces ha repetido la materia

Fuente: El autor

### 6.3.1.2 Competencia

Permite almacenar las competencias que se van a evaluar, para este caso solo se incluyó la competencia lectora, pero cuanta con la flexibilidad de incluir otras.

**Tabla 3** Tabla de competencias

Campo	Tipo	Descripción
-------	------	-------------

Id	Int – Llave primaria	Identificador de la competencia, en el caso en que se evaluen más competencias
Competencia	varchar	Competencia
Descripción	varchar	Descripción de la competencia a evaluar

Fuente: El autor

### 6.3.1.3 Niveles

Permite almacenar información sobre los niveles en los que se evaluara la competencia, está relacionada con la tabla de competencias

**Tabla 4** Tabla de niveles

Campo	Tipo	Descripción
Id	int	Identificador único para el nivel
Nivel	varchar	Nivel de clasificación de las competencias
Descripción	text	Descripción de las habilidades asociadas al nivel
Limite_inferior	int	Puntaje mínimo para clasificar en el nivel
competencia	int	Llave foránea a competencia

Fuente: El autor

### 6.3.1.4 Tareas

Almacena los estímulos que permitirán evaluar, en conjunto con las preguntas, el nivel de habilidad en cada competencia.

**Tabla 5 Tabla de tareas**

Campo	Tipo	Descripción
Id	int	Identificador único de la tarea
Titulo	Varchar	Titulo asociado a la tarea
Texto	Text	Texto asociado a la tarea
Competencia	Int	Competencia a la que pertenece la tarea, llave foránea

Fuente: El autor

### 6.3.1.5 Preguntas

Almacena las preguntas, sus posibles respuestas, la respuesta correcta y el puntaje por pregunta

**Tabla 6 Tabla de preguntas**

Campo	Tipo	Descripción
Id	Int	Identificador único de la pregunta
Pregunta	varchar	Texto de la pregunta
A	Varchar	Opción de respuesta a
B	Varchar	Opción de respuesta b
C	Varchar	Opción de respuesta c
D	Varchar	Opción de respuesta d
Puntaje	Int	Nivel de desempeño con el que está asociado la pregunta
Correcta	Enum (a,b,c,d)	Cuál es la opción correcta

Tarea	Int	A que tarea pertenece (Llave foránea)
-------	-----	---------------------------------------

Fuente: El autor

### 6.3.1.6 Resultados

Es donde se guardan las respuestas a las preguntas por estudiantes

Campo	Tipo	Descripción
Id	int	Identificador único del campo
Estudiante	int	Llave foránea a estudiante que presenta la prueba
Pregunta	int	Pregunta a la que pertenece la respuesta, llave foránea
Respuesta	varchar	Respuesta seleccionada por el estudiante

Fuente: El autor

### 6.3.1.7 Nivel estudiante

Almacena cual es el nivel del estudiante luego de hacer la prueba

**Tabla 7 Tabla Nivel estudiante**

Campo	Tipo	Descripción
Id	Int	Identificador único
Estudiante	Int	Quien presenta la prueba
Competencia	Int	Competencia en cuestión
Nivel	duoble	Nivel de la personan que presenta la prueba

Fuente: El autor

## 7 ANÁLISIS DE RESULTADOS

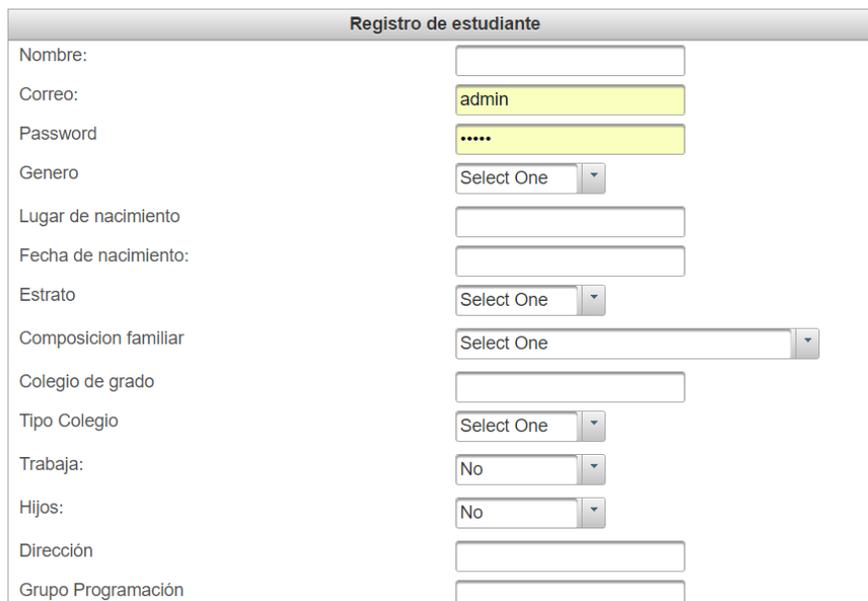
### 7.1 RECOLECCIÓN DE INFORMACIÓN

Para el levantamiento de la información se seleccionaron dos grupos de la materia de “Programación de computadores” de primer semestre de la Universidad Católica de Colombia, con los que se usó una aplicación web diseñada para tal fin. (Ver ANEXO A. MANUAL DE USUARIO “SISTEMA DE LEVANTAMIENTO DE INFORMACIÓN”) y que consta de las siguientes partes.

#### 7.1.1 Contexto

Este formulario tiene como objetivo recoger información personal del estudiante. Las preguntas de contexto solamente permiten seleccionar una respuesta, Figura 11 Cuestionario preguntas de contexto

Figura 11 Cuestionario preguntas de contexto



The image shows a web form titled "Registro de estudiante" with the following fields and values:

Field	Value
Nombre:	
Correo:	admin
Password	.....
Genero	Select One
Lugar de nacimiento	
Fecha de nacimiento:	
Estrato	Select One
Composicion familiar	Select One
Colegio de grado	
Tipo Colegio	Select One
Trabaja:	No
Hijos:	No
Dirección	
Grupo Programación	

Fuente: El autor

#### 7.1.2 Competencia lectora

Esta prueba pretende establecer el nivel de competencia lectora del estudiante. En total son 12 preguntas clasificadas en diferentes niveles. Ver capítulo 5.

La prueba está organizada en estímulos o tareas y preguntas, así, el estímulo es como en Figura 12 Ejemplo de estímulo y la pregunta está asociada al mismo como en Figura 13 Ejemplo de pregunta.

**Figura 12 Ejemplo de estímulo**

▾ LA DONACIÓN DE SANGRE

La donación de sangre es esencial.

No existe ningún producto que pueda sustituir la sangre humana completamente.

La donación de sangre es, por lo tanto, irremplazable y esencial para salvar vidas.

En Francia, cada año, 500.000 pacientes se benefician de una transfusión de sangre.

Los instrumentos para sacar la sangre están esterilizados y se usan una sola vez (jeringas, tubos, bolsas).

Donar sangre no representa ningún riesgo.

**Donación de sangre:**

Es la forma de donación más conocida y toma 45 minutos a 1 hora.

Se extrae una bolsa de 450 ml, además de algunas pequeñas muestras sobre las cuales se harán análisis y controles.

- Un hombre puede donar sangre cinco veces al año y una mujer tres veces.
- Los donantes pueden tener entre 18 y 65 años de edad.

Debe haber un intervalo de 8 semanas entre cada donación.

El texto "Anuncio sobre la donación de sangre", ha sido tomado de un sitio web francés. Úsalo para responder las siguientes preguntas.

Fuente: El autor

**Figura 13 Ejemplo de pregunta**

Preguntas

El texto dice: "Los instrumentos para sacar la sangre están esterilizados y se usan una sola vez ..." ¿Por qué se incluye esta información en el texto?

- Para enfatizar que la donación de sangre es esencial.
- Para asegurar que la donación de sangre es segura.
- Para explicar qué uso se le da a la sangre.
- Para dar detalles sobre los análisis y controles.

Fuente: El autor

## 7.2 DATOS RECOLECTADOS

Los datos obtenidos son los siguientes:

### 7.2.1 Genero

El 21.87 % de estudiantes consultados pertenece al género femenino, contra un 78.12 % de género masculino, lo cual es común encontrar en las clases de la facultad de ingeniería. Figura 14 Estudiantes por género

Figura 14 Estudiantes por género

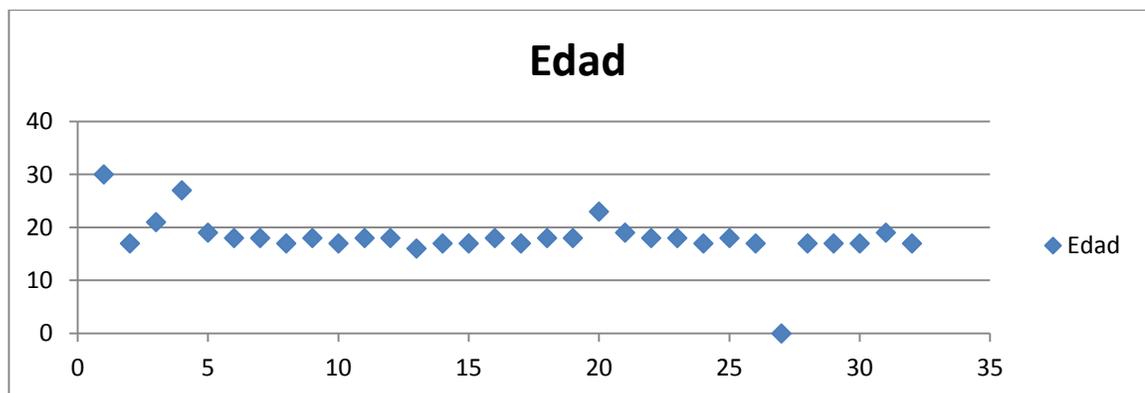


Fuente: El autor

### 7.2.2 Edad

El rango de edad de los estudiantes consultados se mantiene entre los 17 y los 20, salvo algunas excepciones. Figura 15 Edades de estudiantes

Figura 15 Edades de estudiantes

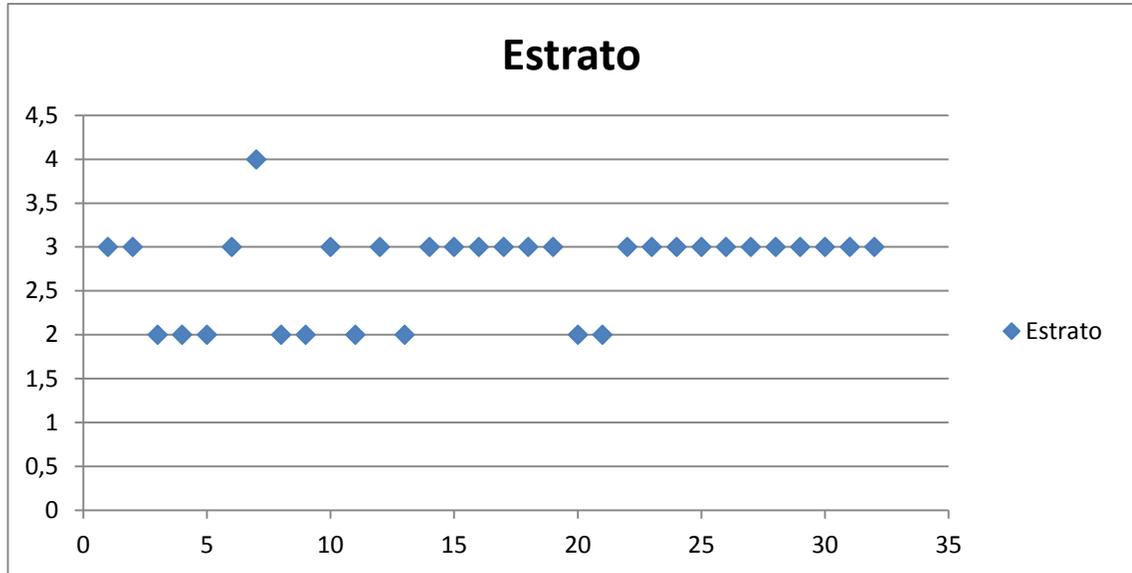


Fuente: El autor

### 7.2.3 Estrato socioeconómico

Los estudiantes participantes se pueden clasificar como clase media, con una mayoría en estrato 3. Figura 16 Estrato de estudiantes

Figura 16 Estrato de estudiantes



Fuente: el autor

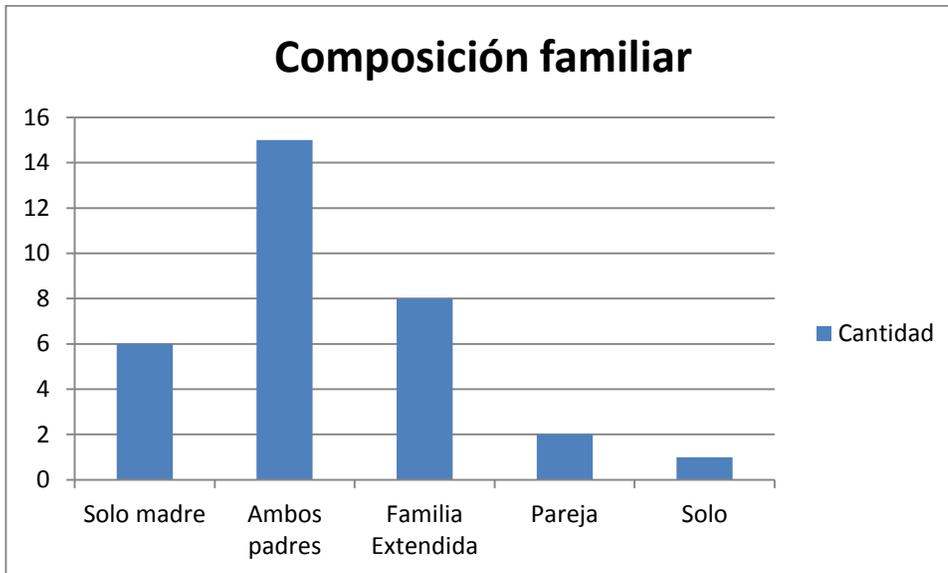
### 7.2.4 Composición familiar

Esta variable se mide con las opciones:

- Solo madre: Hogar conformado con madre cabeza de familia
- Solo padre: Hogar conformado con padre cabeza de familia
- Ambos padres: Hogar con ambos padres presentes
- Familia extendida: Hogar conformado por miembros atípicos, abuelos, tíos, otros familiares
- Pareja: Estudiante ha conformado su propia familia
- Solo: Estudiante vive solo

Los resultados muestran que aunque es más común que los estudiantes vivan con ambos padres, se encuentran una cantidad similar de casos de estudiantes que viven solo con la madre o con otros familiares. Figura 17 Composición Familiar Estudiantes

Figura 17 Composición Familiar Estudiantes

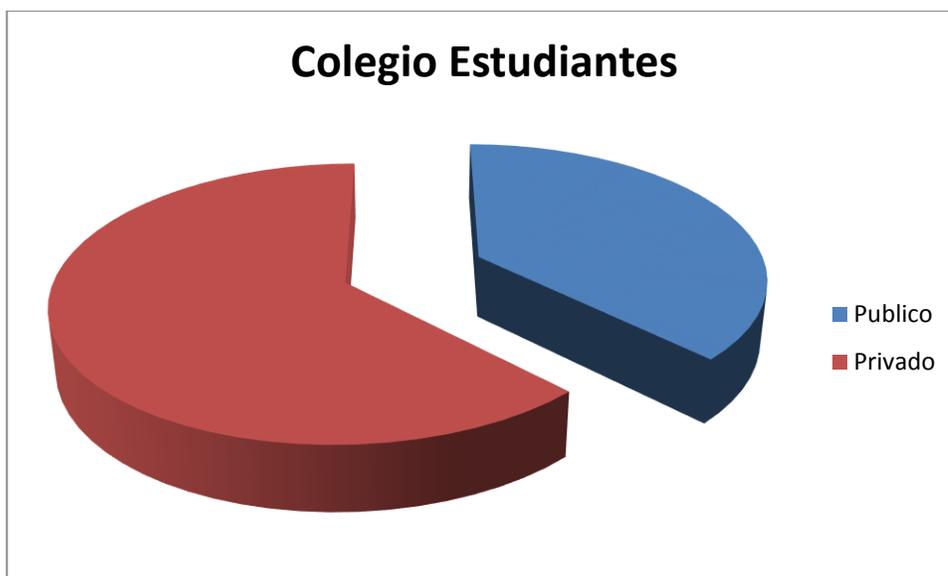


Fuente: El autor

### 7.2.5 Colegio de grado

Aunque se presentan más estudiantes provenientes de colegio privados, hay una porción significativa que viene de colegios públicos. Figura 18 Colegio Estudiantes

Figura 18 Colegio Estudiantes



Fuente: El autor

### 7.2.6 Estudiantes trabajando

La mayoría de los estudiantes no trabaja, lo cual podría corresponder con la edad.  
Figura

Figura 19 Estudiantes que trabajan



Fuente: El autor

### 7.2.7 Estudiantes con hijos

Se analiza que cantidad de estudiante tiene hijos Figura 20 Estudiantes con hijos

Figura 20 Estudiantes con hijos

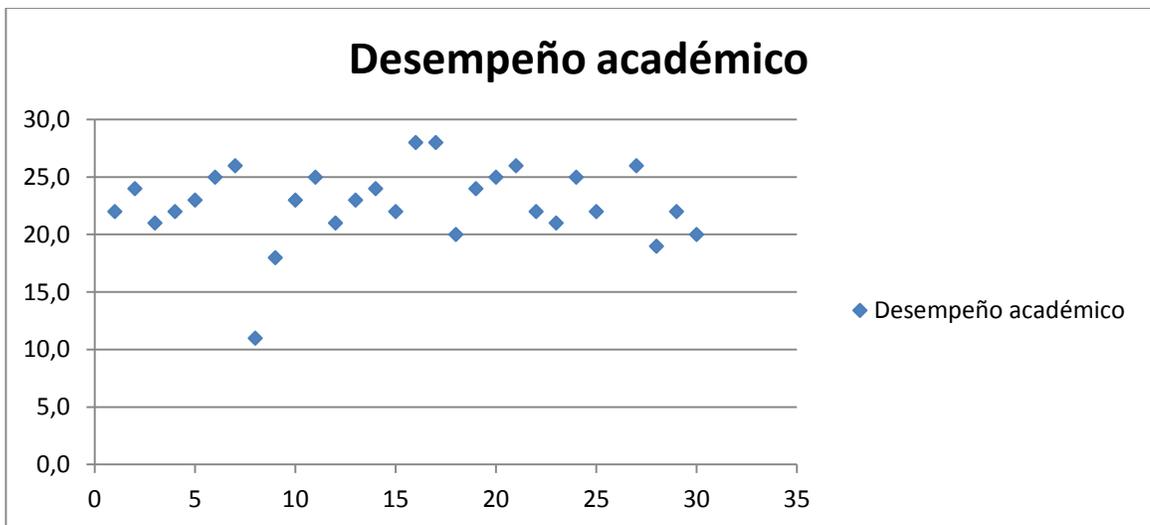


### 7.2.8 Desempeño académico

El desempeño académico de los estudiantes se mide en una escala de 1 a 30 puntos, y corresponde con las notas de primer corte de la asignatura de "Programación de computadores". Figura 21 Desempeño académico

La mayor concentración de notas se encuentra entre 20 y 25 puntos.

Figura 21 Desempeño académico



Fuente: El autor

### 7.3 Clasificación automática de estudiantes

La intención de recolectar y analizar las características de estudiantes es localizar un patrón que permita predecir el comportamiento de dichos estudiantes, para esto se aplican modelos de clasificación que luego de ser entrenados permitan clasificar nuevos estudiantes. Con los datos recolectados y expuestos en 7.2 se organiza un *dataset* donde cada fila representa las características del estudiante, como muestra en Figura 22 Dataset, los datos cualitativos fueron representados por números. Figura 22 Dataset

Figura 22 Dataset

```
Estrato;Composicion familiar;Colegio; trabaja;hijos;Competencia lectora; edad
3;3;1;1;1;0;30
3;3;1;0;0;1;17
2;3;1;0;0;1;21
2;5;0;0;1;1;27
2;1;1;1;0;2;19
```

Fuente: el autor

Para aplicar los algoritmos se usa la biblioteca de para Python llamada *Scikit learn*, que permite la implementación de varios algoritmos de de machine learning.<sup>28</sup>

Luego de leer el dataset se escogen posiciones aleatorias para seleccionar el 30% como conjunto de entrenamiento y el restante 70% como pruebas, la información se guarda en dos arreglos  $X$  de características y  $y$  de etiquetas. El conjunto de entrenamiento se llama training $X$  y las etiquetas del mismo están en training $y$ . Esto es común a los tres modelos a implementar.

#### 7.3.1 K Nearest neighbors

Cando se tiene diferenciado el grupo de pruebas y testing se instancia el clasificador y se entrena con el grupo de entrenamiento, como se ve en la Figura 23 Entrenamiento del modelo de KNeighborsClassifier.

Luego de lo cual se aplica el modelo de predicción al grupo de prueba y se obtienen la clasificación resultante del modelo en un arreglo como se ve en Figura 24 Predicción del modelo

---

<sup>28</sup> Scikit-learn. <http://scikit-learn.org/stable/>

**Figura 23 Entrenamiento del modelo de KNeighborsClassifier**

```
clf = neighbors.KNeighborsClassifier(n_neighbors)

Automatically created module for IPython interactive environment

In [3]: clf.fit(X,y)

Out[3]: KNeighborsClassifier(algorithm='auto', leaf_size=30, metric='minkowski',
                             metric_params=None, n_jobs=1, n_neighbors=4, p=2,
                             weights='uniform')
```

Fuente: El autor

**Figura 24 Predicción del modelo**

```
: predicty=clf.predict(trainingX)
predicty

: array(['2', '3', '3', '3', '2', '4', '2', '2', '3', '3', '3', '2', '1',
        '2', '1', '3', '4', '2', '4', '4', '2', '3', '3', '3', '3', '1'],
       dtype='|S1')
```

Fuente: El autor

Para medir el desempeño del algoritmo se aplica a una matriz de confusión a la que se le pasa el arreglo de clases correspondientes al grupo de pruebas y las clases predichas por el modelo, como en la Figura 25 Inicialización de la matriz de confusión

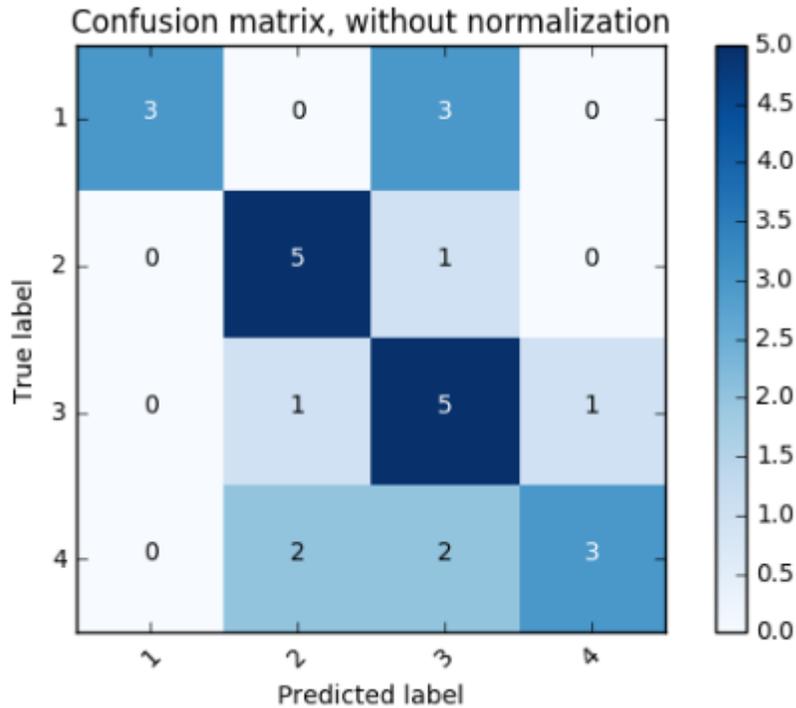
**Figura 25 Inicialización de la matriz de confusión**

```
In [6]: confusion=confusion_matrix(trainingy, predicty)
```

Fuente: El autor

El resultado de la matriz de confusión muestra en las filas horizontales los casos pertenecientes a cada etiqueta, mientras las verticales muestran las clasificaciones dadas por el modelo para los casos, como se ve en Figura 26 Matriz de confusión Neighbors Classifier. Es así como para la etiqueta 1 se tenían 6 casos y 3 fueron clasificados correctamente por el algoritmo clasificador. En el caso dela etiqueta 3 se tienen 7 casos, sin embargo el algoritmo clasifico 1 con la etiqueta 2, 1 con la etiqueta 4 y 5 correctamente.

Figura 26 Matriz de confusión Neighbors Classifier



Fuente: El autor

El reporte de clasificación de la Figura 27 Reporte de clasificación KNeighbors Classifier, muestra que en promedio la precisión, que corresponde la razón entre los casos clasificados correctamente y los casos clasificados por clase es de 0.7, y el recall, que hace referencia a la razón entre los casos clasificados en una etiqueta y los que realmente pertenecen a la etiqueta es de 0.62

La medida de F1-score, que es el peso promedio de las medidas de precisión y recall y que en promedio arroja un 0.62.

Support es el número de casos en cada etiqueta para la clasificación real.

Esto indica que con los casos de prueba se obtiene una capacidad de predicción considerable, con solo 11 casos de entrenamiento.

Figura 27 Reporte de clasificación KNeighbors Classifier

```
: target_names = ['Bajo', 'Medio', 'Alto', 'Superior']  
print(classification_report(trainingy, predicty, target_names=target_names))
```

	precision	recall	f1-score	support
Bajo	1.00	0.50	0.67	6
Medio	0.62	0.83	0.71	6
Alto	0.45	0.71	0.56	7
Superior	0.75	0.43	0.55	7
avg / total	0.70	0.62	0.62	26

Fuente: El autor

### 7.3.2 Random Forest

El procedimiento para este algoritmo es similar que para 7.3.1 K Neares neighbor, en primer lugar se instancia el algoritmo y se entrena con el conjunto de entrenamiento, como en Figura 28 Entrenamiento para Random Forest.

Figura 28 Entrenamiento para Random Forest

```
: clf = RandomForestClassifier(n_estimators=10)  
clf.fit(X, y)  
  
: RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',  
max_depth=None, max_features='auto', max_leaf_nodes=None,  
min_impurity_split=1e-07, min_samples_leaf=1,  
min_samples_split=2, min_weight_fraction_leaf=0.0,  
n_estimators=10, n_jobs=1, oob_score=False, random_state=None,  
verbose=0, warm_start=False)
```

Fuente: El Autor

Luego se realiza la predicción para el conjunto de prueba, Figura 29 Predicción para Random Forest.

Figura 29 Predicción para Random Forest

```
predicty = clf.predict(trainingX)
predicty
array(['3', '2', '2', '3', '3', '4', '2', '4', '2', '2', '2', '3', '2',
       '1', '3', '1', '4', '4', '2', '4', '4', '2', '2', '1', '1', '1'],
      dtype='|S1')
```

Fuente: El autor

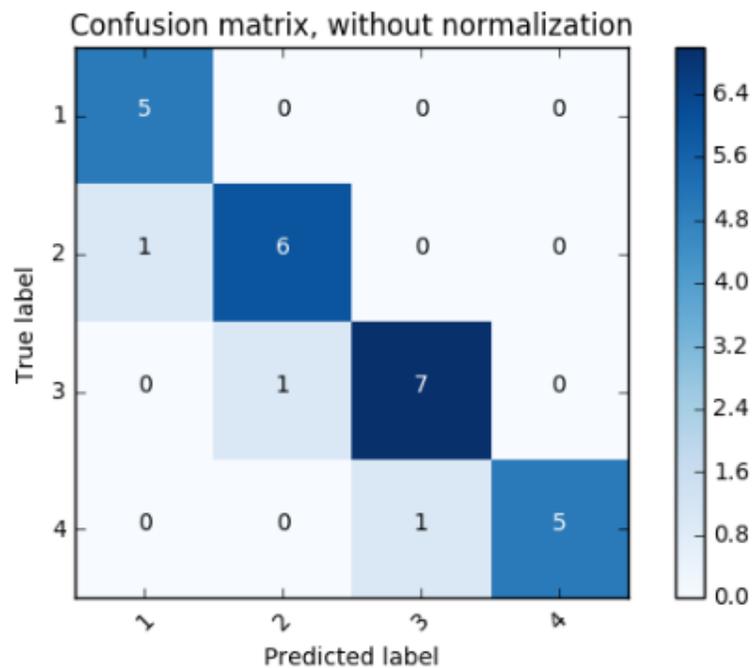
La matriz de confusión para Random Forest está en Figura 30 Matriz de confusión para Random Forest, muestra la predicción para este caso, que mejora el de Nearest neighbors

Figura 30 Matriz de confusión para Random Forest

```
In [7]: class_names=['1', '2', '3', '4']
# Plot non-normalized confusion matrix
plt.figure()
plot_confusion_matrix(confusion, classes=class_names,
                      title='Confusion matrix, without normalization')
plt.show()
```

Confusion matrix, without normalization

```
[[5 0 0 0]
 [1 6 0 0]
 [0 1 7 0]
 [0 0 1 5]]
```



Fuente: El autor

El reporte de clasificación se ve en Figura 31 Reporte de clasificación para Random Forest, que muestra lo atinado del modelo en este caso.

**Figura 31 Reporte de clasificación para Random Forest**

```
target_names = ['Bajo', 'Medio', 'Alto', 'Superior']
print(classification_report(trainingy, predicty, target_names=target_names))
```

	precision	recall	f1-score	support
Bajo	0.83	1.00	0.91	5
Medio	0.86	0.86	0.86	7
Alto	0.88	0.88	0.88	8
Superior	1.00	0.83	0.91	6
avg / total	0.89	0.88	0.88	26

Fuente: El autor

### 7.3.3 Support Vector Classifier

Para este algoritmo se tiene la instanciación y entrenamiento en Figura 32 Entrenamiento SVC

**Figura 32 Entrenamiento SVC**

```
clf=SVC()
clf.fit(X,y)
```

Automatically created module for IPython interactive environment

```
SVC(C=1.0, cache_size=200, class_weight=None, coef0=0.0,
    decision_function_shape=None, degree=3, gamma='auto', kernel='rbf',
    max_iter=-1, probability=False, random_state=None, shrinking=True,
    tol=0.001, verbose=False)
```

Fuente: El autor

Luego, se realiza la predicción sobre el conjunto de prueba. Figura 33 Predicción para SVC

Figura 33 Predicción para SVC

```
: predicty=clf.predict(trainingX)
predicty
: array(['3', '3', '3', '2', '2', '4', '2', '2', '2', '3', '4', '2', '1',
        '3', '1', '4', '3', '4', '2', '4', '2', '2', '1', '1', '1'],
        dtype='|S1')
```

Fuete: El autor

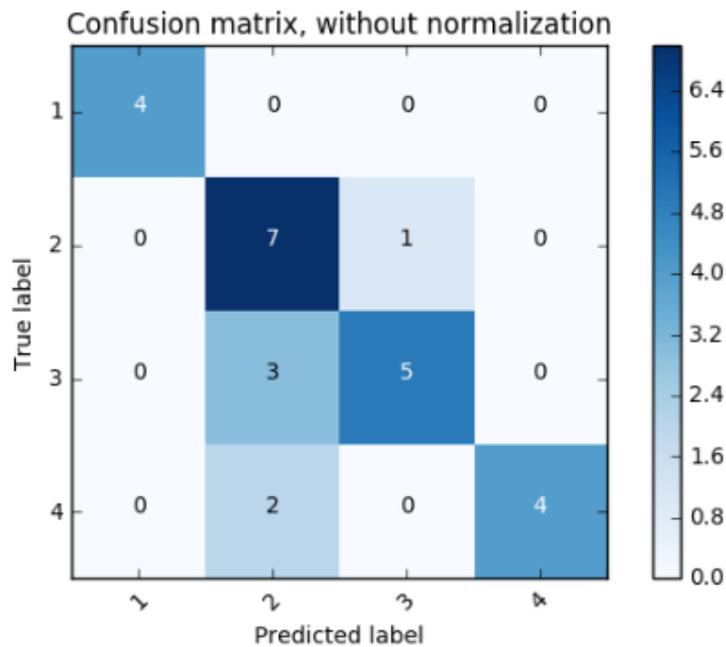
La matriz de confusión para este caso se muestra en Figura 34 Matriz de confusión para SVC

Figura 34 Matriz de confusión para SVC

```
class_names=['1', '2', '3','4']
# Plot non-normalized confusion matrix
plt.figure()
plot_confusion_matrix(confusion, classes=class_names,
                      title='Confusion matrix, without normalization')
plt.show()
```

Confusion matrix, without normalization

```
[[4 0 0 0]
 [0 7 1 0]
 [0 3 5 0]
 [0 2 0 4]]
```



Fuete: El autor

Luego se puede observar el reporte de clasificación para este algoritmo en Figura 35 Reporte de clasificación para SVC

Figura 35 Reporte de clasificación para SVC

```
: target_names = ['Bajo', 'Medio', 'Alto', 'Superior']  
print(classification_report(trainingy, predicty, target_names=target_names))
```

	precision	recall	f1-score	support
Bajo	1.00	1.00	1.00	4
Medio	0.58	0.88	0.70	8
Alto	0.83	0.62	0.71	8
Superior	1.00	0.67	0.80	6
avg / total	0.82	0.77	0.77	26

Fuente: El autor

## 8 DISEÑO Y CONSTRUCCIÓN DEL TUTOR INTELIGENTE

### 8.1 DISEÑO DEL MÓDULO DEL DOMINIO DEL CONOCIMIENTO

Para manejar la incertidumbre asociada a las rutas de aprendizaje se propone usar Redes Bayesianas, (BN)<sup>29</sup>, las cuales usan la teoría de la probabilidad como herramienta para manejar situaciones como esta. BN ha sido aplicado con éxito en una amplia variedad de problemas, como los sistemas de propulsión a chorro de transbordadores espaciales de la NASA<sup>30</sup> y en el Asistente de Office, el paquete de Microsoft.

La BN que se propone modela la probabilidad de independencia condicional entre los conceptos, los cuales serán extraídos de un curso de programación básica. Cada concepto es representado como un nodo en el grafo de la red bayesiana. Se agrega una conexión, o arista, de un nodo a otro si el concepto predecesor es prerequisite para entender el concepto sucesor. De esta forma la BN puede ser construida manualmente con la ayuda de un texto guía del tema<sup>31</sup>. Así, por ejemplo, para entender el concepto de un ciclo For, primero es necesario entender los conceptos de asignación de variables, operadores relacionales y operadores incrementales. La representación de esta relación puede hacerse a través de una red bayesiana como la de la Figura 36 Subgrafo para el concepto de ciclo For.

---

<sup>29</sup> PEARL, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

<sup>30</sup> HORVITZ, Eric; BARRY, Matthew. Display of information for time-critical decision making. En *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995. p. 296-305.

<sup>31</sup> BUTZ, C. J.; HUA, S.; MAGUIRE, R. B. Bits: a bayesian intelligent tutoring system for computer programming. En *Western Canadian Conference on Computing Education*. 2004.

Figura 36 Subgrafo para el concepto de ciclo For



Fuente: Bits: a Bayesian Intelligent Tutoring System For Computer Programming

Este caso es solamente una parte de lo que sería el grafo completo.

Basándose en este grafo es posible construir la probabilidad condicionalmente independiente para el concepto del ciclo FOR, dado los nodos padres. Dónde:

$a =$  Concepto del ciclo For

$P_i =$  El conjunto de los coceptos padres de  $a$

La probabilidad condicional, que establece si se conoce o no el concepto del ciclo For, dado si se conocen o no conocen los conceptos padres, queda así:

$$p(a|P_i) = p(a|Asignación, Operadores relacionales, Operadores incrementales y decrementales)$$

Cuya probabilidad total se puede representar así:

Tabla 8 Probabilidad total para el concepto de ciclo For

Nodos padres			Ciclo For	
Asignación de	Operadores Relacionales	Operadores incrementales y decrementales	Sabe	No sabe

Variables				
Sabe	Sabe	Sabe	0.75	0.25
		No sabe	0.39	0.61
	No sabe	Sabe	0.50	0.50
		No sabe	0.22	0.78
No sabe	Sabe	Sabe	0.50	0.50
		No sabe	0.29	0.71
	No sabe	Sabe	0.40	0.60
		No sabe	0.15	0.85

Fuente: Bits: a Bayesian Intelligent Tutoring System For Computer Programming

### 8.1.1 Red bayesiana para el curso de programación de computadores

Manteniendo la intención de tomar el curso de programación de computadores de la Universidad Católica de Colombia, se propone un modelo de red bayesiana para la primera parte del curso.

Este modelo bayesiano es ideal y asume equiprobabilidad para todos los nodos, para probar el concepto, es decir, la probabilidad de saber (Know) y no saber (Unknow) es la misma para los nodos con jerarquías simples, es decir aquellos que solo tienen un padre. Los nodos que tienen varios padres se plantean pensando en que existe más probabilidad de dominar el concepto en cuanto se dominen sus conceptos padres y menos cuando no sea así.

Es importante impulsar a que futuros trabajos prueben estas relaciones para obtener redes que se adapten mejor a la realidad.

La red propuesta incluye los siguientes conceptos representados en nodos como se especifica en la Tabla 9 Conceptos para red bayesiana

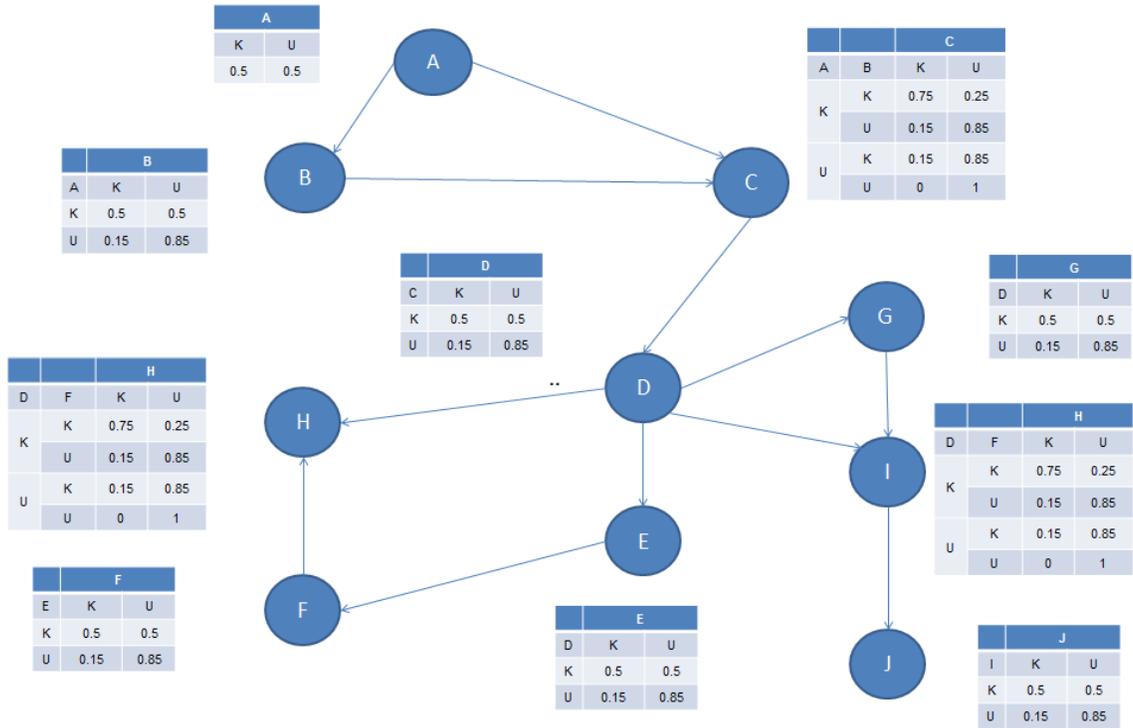
**Tabla 9 Conceptos para red bayesiana**

Concepto	Nodo
Problemas de una y dos dimensiones	A
Solución de problemas dinámicos	B
Algoritmos	C
Tipos de datos	D
Declaración de variables	E
Asignación de variables	F
Expresiones aritmético lógicas	G
Entrada y salida de información	H
Sentencias de decisión	I
Sentencias repetitivas	J

Fuente: El autor

La red y sus probabilidades quedan planteadas como en la Figura 37 Red bayesiana propuesta

Figura 37 Red bayesiana propuesta



Fuente: El autor

Para la implementación de la red se ha usado la biblioteca Jayes, desarrollada por Michael Kutschke de la maestría en Ciencias de la computación de la Universidad Technische Universität Darmstadt. La biblioteca hace parte del proyecto Eclipse Code Recommenders.<sup>32</sup>

La implementación de la red en Jayes es muy sencilla y se divide en la construcción de la red y en la inferencia.

La construcción de la red se muestra en la Figura 38 Ejemplo de construcción de la red en Jayes, donde en primer lugar se crea la red, luego el nodo inicial A, que no tiene nodos padres y que se asume con probabilidades iguales para sus dos estados. Luego se crea el nodo B, donde se puede observar que se establece A como su nodo padre y las consecuentes relaciones según cada estado de A para cada estado de B; en este punto, si el concepto padre es manejado se asumen probabilidades iguales para cada estado, de lo contrario se asume que existe

<sup>32</sup> KUTSCHKE, Michael. Code trails. An introduction to Bayesian Networks with Jayes.

menos probabilidad de que se maneje el concepto si no se maneja el concepto padre.

Figura 38 Ejemplo de construcción de la red en Jayes

```
//Se crea la red
BayesNet net = new BayesNet();

//Se crean los nodos
// Problemas de una y dos dimensiones
BayesNode a = net.createNode("a");
a.addOutcomes("know", "unknown");
a.setProbabilities(0.5,0.5);

//Solucion de problemas dinamicos
BayesNode b = net.createNode("b");
b.addOutcomes("know", "unknown");
b.setParents(Arrays.asList(a));
b.setProbabilities(
    0.5, 0.5, // a == know
    0.15, 0.85// a == unknow
);
```

Fuente: El autor

## 8.2 DISEÑO DEL MÓDULO DEL ESTUDIANTE

Para el alcance de este proyecto, la variable que se tendrá en cuenta para modelar al estudiante es su nivel de conocimiento. Un modelo ampliamente usado para este propósito es el de Rastreo Bayesiano de Conocimiento (*Bayesian Knowledge Tracing*)<sup>33</sup>, que se caracteriza por su precisión, fácil interpretación y habilidad para inferir el conocimiento del estudiante. Este modelo tiene en cuenta diferentes circunstancias que se pueden encontrar cuando el estudiante responde preguntas y que se deben tener en cuenta antes de determinar si el estudiante sabe o no un concepto, como son la probabilidad de que el estudiante sepa la habilidad antes de intentar el primer test, la probabilidad de que el estudiante responda bien el test pero que aun así no sepa la habilidad, y la probabilidad de

---

<sup>33</sup> FAULHABER, Arndt; MELIS, Erica. An Efficient Student Model Based on Student Performance and Metadata. En *ECAI*. 2008. p. 276-280.

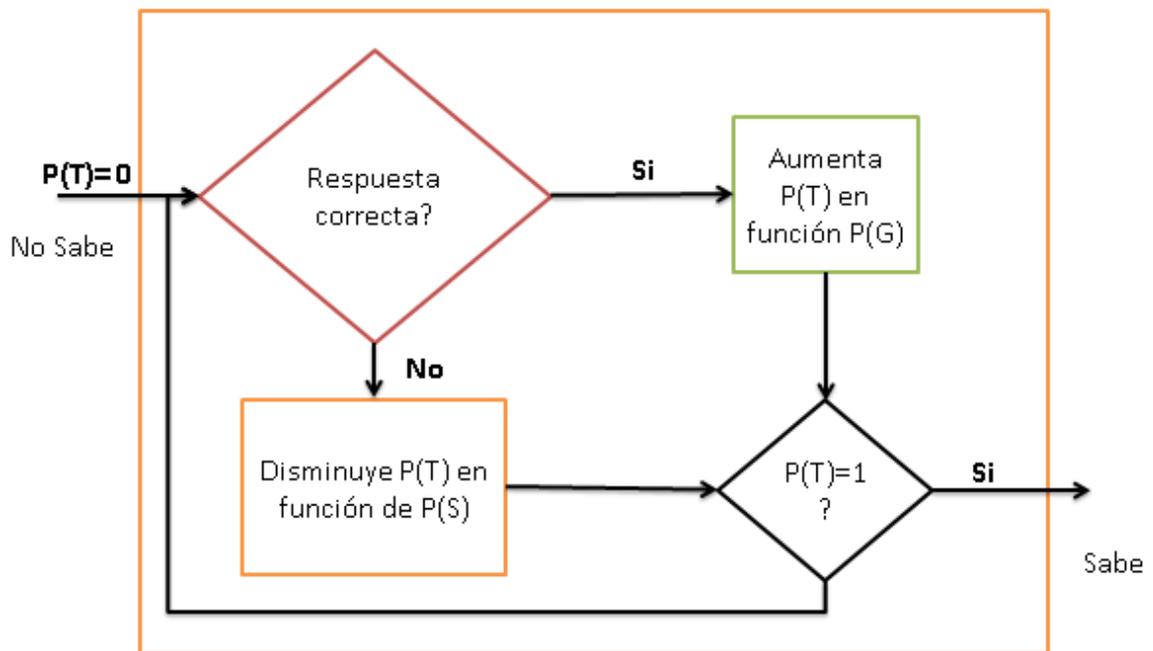
que responda incorrectamente a pesar de que sabe la habilidad<sup>34</sup> y de esta forma garantizar que un estudiante no pasará a estado de conocer un concepto por azar.

Así, analizando un ejemplo sencillo, cuando un estudiante comienza a usar el tutor, el valor de su probabilidad de conocer un concepto es 0,  $P(L_0)=0$ , es probable que el estudiante tenga un conocimiento previo, pero no hay forma de tener una certeza, así que se asume en cero. A continuación el estudiante comienza a responder preguntas de los test, y si él responde correctamente, la probabilidad de que sepa el concepto se incrementa, así puede volverse un 0,2, luego un 0.5 y después un 0.7. Si el estudiante falla en la pregunta la probabilidad decrece. En ambas situaciones hay una incertidumbre, si falla la pregunta puede que aun así la sepa y si la acierta, puede que haya adivinado,  $P(G)$ ; así, si falla luego de haber contestado correctamente dos veces, se asume que fue un desliz  $P(S)$  y su probabilidad de conocer el concepto disminuye en menor cantidad. El propósito es que el estudiante tenga las probabilidades de cada concepto en 1, lo cual quiere decir que se tienen una seguridad del 100% de que el estudiante domina el concepto, que será el valor  $P(T)$ , que quiere decir que el estudiante fue de un estado de no aprendizaje a un estado de aprendizaje, como ilustra la Figura 39 Rastreo de conocimiento.

---

<sup>34</sup> HAWKINS, William J.; HEFFERNAN, Neil T.; BAKER, Ryan SJD. Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. En *International Conference on Intelligent Tutoring Systems*. Springer International Publishing, 2014. p. 150-155.

Figura 39 Rastreo de conocimiento



Fuente: El autor

La labor de determinar cómo se relaciona una pregunta con uno o varios conceptos, de nuevo recaerá en expertos y estará basada en los conceptos resultantes de plantear el dominio del conocimiento.

Los puntos a favor de construir estas dos partes correctamente, de tener los conceptos bien modelados y tener preguntas por cada concepto se reflejan en dos beneficios para el estudiante. El primero, que se tendrá un tutor que no permitirá al estudiante continuar si no demuestra que tiene la habilidad y el segundo, es que luego de que el estudiante ha demostrado tener la habilidad, no se insiste en preguntas repetitivas, lo que hará el aprendizaje más eficiente, pues el estudiante no se entrenará de más en una habilidad.

## 9 CONCLUSIONES

- La selección de la teoría bayesiana como técnica propuesta para el desarrollo del tutor ofrece una herramienta para establecer la ruta de aprendizaje del estudiante tomando como base el temario sugerido por un asesor.
- El diseño de la red bayesiana basada en el contenido del curso de programación de computadores, en conjunto con las conclusiones obtenidas del proceso de modelado de características; sirven como plataforma para el desarrollo de un tutor inteligente más especializado.
- Las características seleccionadas buscan un criterio cuantitativo y cualitativo para obtener información del estudiante que sirva para diferenciarlo y buscar patrones de cómo influyen en su desempeño en programación de computadores
- Se encontró que al separar las características entre contexto y competencia se puede analizar el estudiante desde la influencia que tiene sobre él su entorno, así como obtener un perfil de sus habilidades y conocimientos.
- Se diseñó una base de datos que permite almacenar estas características y que facilita la interpretación y el análisis de los resultados.
- Se encontró un clasificador automático para los estudiantes según su contexto y su competencia basándose en Aprendizaje de Máquina para establecer una relación con su desempeño académico.
- Tomando como punto de partida el diseño propuesto, se procedió a hacer una evaluación de la precisión de los modelos empleados en términos de entradas y resultados obtenidos, mostrando que los resultados obtenidos muestran una relación prometedora entre las características y el desempeño.

## 10 RECOMENDACIONES

- La propuesta de la red bayesiana del tutor inteligente asume que todos los nodos son equiprobables, lo cual posteriormente debe ser sometido a ajustes.
- En estudios posteriores se deberían incluir más variables, tanto en el diseño del tutor, lo cual lo hará más empático; como en el análisis de los modelos de clasificación, lo que arrojará resultados más precisos y útiles.
- Durante la recolección de características se observó que algunos estudiantes no muestran un dominio básico de herramientas como el computador e Internet, factor que puede interferir en su desempeño en la materia, se sugiere incluir esta característica en futuros experimentos.
- La medición de la comprensión de lectura en los estudiantes se aprecia obstaculizada por la actitud de los estudiantes ante la lectura, lo cual puede indicar, inicialmente la magnitud de sus falencias en esta competencia, pero de lo cual se concluye que los datos en este punto no son totalmente precisos. Para futuros experimentos se sugiere mejorar la medición de esta variable.
- Incluir en futuros diseños del tutor un sistema de red bayesiana dinámico y configurable permitirá cubrir más temas y más población estudiantil manteniendo el mismo concepto.
- El modelo se debe probar con un grupo más grande, así se tendrá un grupo de entrenamiento y prueba que permita obtener mejores resultados de los algoritmos de clasificación y llegar a mejores conclusiones
- Se debería experimentar con otros clasificadores que permitan obtener predicciones más precisas.

## BIBLIOGRAFÍA

AHUJA, Neelu Jyothi; SILLE, Roohi. A critical review of development of intelligent tutoring systems: Retrospect, present and prospect. *International Journal of Computer Science Issues*, 2013, vol. 10, no 4, p. 39-48.

ANDERSON, John R. ; SKWARECKI, Edward. The automated tutoring of introductory computer programming. *Communications of the ACM*, 1986, vol. 29, no 9, p. 842-849.

BEN-ARI, Mordechai. Constructivism in computer science education. *Journal of Computers in Mathematics and Science Teaching*, 2001, vol. 20, no 1, p. 45-74.

BISHOP, Christopher M. Pattern recognition. *Machine Learning*, 2006, vol. 128.

BUTZ, C. J.; HUA, S.; MAGUIRE, R. B. Bits: a bayesian intelligent tutoring system for computer programming. En *Western Canadian Conference on Computing Education*. 2004.

BBC ACTIVE. Why you should use a Virtual Learning Environmen [en línea]. [citado 29 de octubre, 2015]. Disponible en Internet: <http://goo.gl/d8YjDP>

CHEN, Xin; BARNETT, Deborah R.; STEPHENS, Casheena. Fad or future: The advantages and challenges of massive open online courses (MOOCs). En *research-to practice conference in adult and higher education*. 2013. p. 20-21.

CONCEPT TO CLASSROOM. Costructivism as a Paradigm for teaching and learning <http://www.thirteen.org/edonline/concept2class/constructivism/>

COURSERA. MACHINE LEARNING. <https://es.coursera.org/learn/machine-learning>

DEFINICIÓN.DE. Definición de metacognición. <http://definicion.de/metacognicion/>

Department of Statistics, University of California, Berkeley. Random forest

DEPARTMENT OF COMPUTER SCIENCE PRINCETON UNIVERSITY. What is machine learning?, theoretical machine learning [en línea], Rob Schapire. [citado 29 de octubre, 2015]. Disponible en Internet: <http://goo.gl/WHZ1jq>

Department de Computer Science. University of Regina. Confussion Matrix. [http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion\\_matrix/confusion\\_matrix.html](http://www2.cs.uregina.ca/~dbd/cs831/notes/confusion_matrix/confusion_matrix.html)

EL CONFIDENCIAL. ¿Colegios públicos o privados? El informe PISA analiza cuáles son mejores. <http://tinyurl.com/ptdlrlz>

EL ESPECTADOR. “Hay déficit de 15.000 ingenieros” para industria TIC: Mintic [en línea]. Colombia: Economía [citado >27 de octubre, 2015]. Disponible en Internet: <http://goo.gl/awS10S>

ELKAN, Charles. Nearest neighbor classification. *elkan@cs.ucsd.edu*, *11 January*, 2011, vol. 11, p. 3.

FAULHABER, Arndt; MELIS, Erica. An Efficient Student Model Based on Student Performance and Metadata. En *ECAI*. 2008. p. 276-280.

HAMID, Oussama H.; ALAIWY, FatemahH; HUSSIEN, Intisar O. A Bayesian Network Model for More Natural Intelligent Tutoring Systems. *International Journal of Enhanced Research in Science, Technology & Engineering*, ISSN, 2015, p. 2319-7463.

HAWKINS, William J.; HEFFERNAN, Neil T.; BAKER, Ryan SJD. Learning bayesian knowledge tracing parameters with a knowledge heuristic and empirical probabilities. En *International Conference on Intelligent Tutoring Systems*. Springer International Publishing, 2014. p. 150-155.

HORVITZ, Eric; BARRY, Matthew. Display of information for time-critical decision making. En *Proceedings of the Eleventh conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1995. p. 296-305.

HORVITZ, Eric, et al. The Lumiere project: Bayesian user modeling for inferring the goals and needs of software users. En *Proceedings of the Fourteenth conference on Uncertainty in artificial intelligence*. Morgan Kaufmann Publishers Inc., 1998. p. 256-265.

IMAGINE BLOG. Bayesian Knowledge Tracing [en línea]. Matthew W. Johnson. Julio 2012. [citado 29 de Octubre de 2015]. Disponible en Internet <http://goo.gl/rj10sF>

Imperial Colege London. Introduction to Neural Networks. <http://tinyurl.com/zf8alvk>

INSTITUCIÓN NACIONAL DE ASTROFÍSICA, ÓPTICA Y ELECTRÓNICA. Luis Enrique Sucar. <https://ccc.inaoep.mx/~esucar/Clases-mgp/caprb.pdf>

KUTSCHKE, Michael. Code trails. An introduction to Bayesian Networks with Jayes.

LANGLEY, Pat. *Elements of machine learning*. Morgan Kaufmann, 1996.

MINISTERIO TIC COLOMBIA EN FACEBOOK. [En línea]. Colombia: Ministerio TIC Colombia [citado 27 de octubre, 2015]. Disponible en Internet: <https://goo.gl/W6Es6c>

NAKABAYASHI, Kiyoshi, et al. An intelligent tutoring system on World-Wide Web: Towards an integrated learning environment on a distributed hypermedia. En *Proc. ED-MEDIA*. 1995.

NWANA, Hyacinth S. Intelligent tutoring systems: an overview. *Artificial Intelligence Review*, 1990, vol. 4, no 4, p. 251-277.

OCDE. Evaluación de las competencias lectoras para el siglo XXI. Pág 21

PEARL, Judea. *Probabilistic reasoning in intelligent systems: networks of plausible inference*. Morgan Kaufmann, 2014.

PILLAY, Nelishia. Developing intelligent programming tutors for novice programmers. *ACM SIGCSE Bulletin*, 2003, vol. 35, no 2, p. 78-82.

RADAR. Learning programming at scale [en línea]. Philip Gou. [citado 29 de Octubre de 2015] Disponible en Internet <http://goo.gl/jEVJuO>

REVISTA SEMANA. ¿Y dónde están los ingenieros? [En línea]. Colombia: Tecnología [citado 27 de octubre, 2015]. Disponible en Internet: <http://goo.gl/rTE5qh>

SAYLOR.ORG. Supervised learning [en línea]. [citado 29 de Octubre de 2015] Disponible en Internet <http://www.saylor.org/site/wp-content/uploads/2011/11/CS405-6.2.1.2-WIKIPEDIA.pdf>

SELF, John. The defining characteristics of intelligent tutoring systems research: ITSs care, precisely. *International Journal of Artificial Intelligence in Education (IJAIED)*, 1998, vol. 10, p. 350-364.

Scikit-learn. <http://scikit-learn.org/stable/>

Scikit learn. Precision-Recall. [http://scikit-learn.org/stable/auto\\_examples/model\\_selection/plot\\_precision\\_recall.html](http://scikit-learn.org/stable/auto_examples/model_selection/plot_precision_recall.html)

SLEEMAN, Derek. The challenges of teaching computer programming. *Communications of the ACM*, 1986, vol. 29, no 9, p. 840-841.

Stanford Encyclopedia of Philosophy. Fuzzy Logic <http://plato.stanford.edu/entries/logic-fuzzy/>

STATHACOPOULOU, R.; MAGOULAS, G. D.; GRIGORIADOU, M. Neural network-based fuzzy modeling of the student in intelligent tutoring systems. En *Neural Networks, 1999. IJCNN'99. International Joint Conference on*. IEEE, 1999. p. 3517-3521

THE MIT ENCYCLOPEDIA OF THE COGNITIVE SCIENCES. Unsupervised Learning. Peter Dayan

UNIVERSITY OF CAMBRIDGE. Unsupervised Learning. Zoubin Ghahramani. Septiembre de 2004.

UNIVERSITY OF WASHINGTON COMPUTER SCIENCE & ENGINEERING. Bayesian networks.

VARGAS, Guiselle María Garbanzo. Factores asociados al rendimiento académico en estudiantes universitarios desde el nivel socioeconómico: Un estudio en la Universidad de Costa Rica. *Revista Electrónica Educare*, 17 (3), 2013, p. 57-87.

WANG, Yutao; HEFFERNAN, Neil T. The student skill model. En *International Conference on Intelligent Tutoring Systems*. Springer Berlin Heidelberg, 2012. p. 399-404.

WONG, S.. K.. Michael ; BUTZ, Cory J.. ; WU, Dan. On the implication problem for probabilistic conditional independency. *IEEE Transactions on Systems, Man, and Cybernetics-Part A: Systems and Humans*, 2000, vol. 30, no 6, p. 785-805.

## **ANEXOS**

### **ANEXO A. MANUAL DE USUARIO “SISTEMA DE LEVANTAMIENTO DE INFORMACIÓN”**

**Manual de usuario “Sistema de levantamiento de información ”**

**Proyecto de grado**

**Diseño e implementación de un tutor inteligente para enseñar a programar**

**Luisa Fernanda Quiroga**

**[lfquiroga04@ucatolica.edu.co](mailto:lfquiroga04@ucatolica.edu.co)**

**Universidad Católica de Colombia**

## 1. Registro de estudiantes:

### 1.1. Página inicial

<http://54.186.137.128:8080/AteneaV1>



Ingreso	
Usuario	<input type="text" value="admin"/>
Clave	<input type="password" value="...."/>
<input type="button" value="Login"/>	
<input type="button" value="Registrar"/>	

Seleccione el botón “Registrar”

### 1.2. Registro

Responda con la mayor sinceridad posible

A continuación se presentan unas preguntas de contexto, las cuales permitirán la posibilidad de obtener nuevos resultados sobre la influencia de ciertas variables en el proceso educativo. Se ruega completar toda la información y con la mayor sensibilidad.

Registro de estudiante	
Nombre:	<input type="text"/>
Correo:	<input type="text" value="admin"/>
Password	<input type="password" value="...."/>
Genero	<input type="text" value="Select One"/>
Lugar de nacimiento	<input type="text"/>
Fecha de nacimiento:	<input type="text"/>
Estrato	<input type="text" value="Select One"/>
Composicion familiar	<input type="text" value="Select One"/>
Colegio de grado	<input type="text"/>
Tipo Colegio	<input type="text" value="Select One"/>
Trabaja:	<input type="text" value="No"/>
Hijos:	<input type="text" value="No"/>
Dirección	<input type="text"/>
Grupo Programación	<input type="text"/>

### 1.3. Guardar

Cuando haya solucionado todos los puntos de click en guardar

Repetente ~

Cuantas veces ha repetido

## 1.4. Error al guardar

Un mensaje como el siguiente indicará campos faltantes

RegistrarEstudiante:j\_idt8: Error de validación: se necesita un valor.  
Debe ser un correo valido de la universidad  
RegistrarEstudiante:Genero: Error de validación: se necesita un valor.  
RegistrarEstudiante:lugarNacimiento: Error de validación: se necesita un valor.  
RegistrarEstudiante:estrato: Error de validación: se necesita un valor.  
RegistrarEstudiante:composicionFamiliar: Error de validación: se necesita un valor.  
RegistrarEstudiante:tipoColegio: Error de validación: se necesita un valor.  
RegistrarEstudiante:j\_idt49: Error de validación: se necesita un valor.  
RegistrarEstudiante:j\_idt51: Error de validación: se necesita un valor.

Registro de estudiante

Luego de guardar será redirigido a la página inicial

## 2. Login



ATENEA

Para aprender como aprendemos



UNIVERSIDAD CATÓLICA  
de Colombia

Ingreso	
Usuario	<input type="text" value="admin"/>
Clave	<input type="password" value="....."/>
Login	
Registrar	

Use las credenciales que acaba de crear

### 2.1. Informe de objetivo

Acepto responder esta prueba y entiendo su objetivo? [Acepto](#)

Acepte para iniciar la prueba

## 3. Prueba

El menu es un acordeon, al tocar cada opción se desplegará un texto que es el cuerpo de la prueba



▶ Trabajo a distancia

▶ Lo que importa es la obra

▶ LA DONACIÓN DE SANGRE

Enviar Test



▼ Trabajo a distancia

#### El camino del futuro

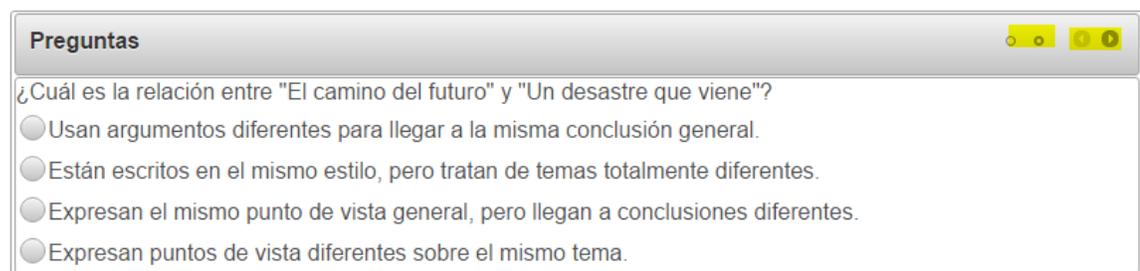
Imagínate lo maravilloso que sería trabajar a distancia\*, desplazándote a través de la carretera electrónica con todo tu trabajo hecho en un computador o por teléfono! Ya no tendrías que viajar en buses o trenes llenos de gente o pasar horas y horas trasladándote desde y hacia el trabajo. Podrías trabajar cuando lo desees; Sólo piensa en todas las oportunidades de trabajo que tendrías!

Mariana

#### Un desastre que viene

Reducir las horas de traslado y reducir el consumo de energía que ello implica es obviamente una buena idea. Pero dicho objetivo debería lograrse mejorando el sistema de transporte público o procurando que los lugares de trabajo estén ubicados cerca de donde vive la gente. La ambiciosa idea de que el trabajo a distancia debería ser parte de la vida de toda la gente sólo lograría que las personas se volvieran cada vez más ensimismadas. ¿Realmente queremos que nuestro sentido de ser

Al final de cada sección encontrará las preguntas a resolver



**Preguntas**

¿Cuál es la relación entre "El camino del futuro" y "Un desastre que viene"?

- Usan argumentos diferentes para llegar a la misma conclusión general.
- Están escritos en el mismo estilo, pero tratan de temas totalmente diferentes.
- Expresan el mismo punto de vista general, pero llegan a conclusiones diferentes.
- Expresan puntos de vista diferentes sobre el mismo tema.

Asegurese de resolver todas las preguntas

Al guardar

---

**Enviar Test**

Si aun hay preguntas sin resolver, verá algo como esto



**Debes responder todas las preguntas  
Debes responder todas las preguntas**

