Luminus

Escuela Superior de Cómputo del IPN Escuela Superior de Cómputo, IPN

16 de octubre de 2018





Índice general

1.	Capitulo 1	1
	1.1. Introduccion	1
	1.2. Introduccion	2
	1.3. Introduccion	3
	1.4. Introduccion	4
	1.5. Introduccion	5
	1.6. Introduccion	
	1.7. Introduccion	7
	1.8. Introduccion	
	1.9. Introduccion	
	1.10. Introduccion	10
2	Capitulo 2	13
۷.	2.1. Marco Teórico	
	2.1.1. Big Data	
	2.1.2. Minería de datos	
	2.1.3. Árboles de decisión	
	2.1.4. ID3 (Iterative Dichotomiser 3)	
	2.1.5. C4.5	
	2.1.6. Spark	
	2.1.7. Hadoop	
	2.1	
3.	Capitulo 3	23
	3.1. Instalación de la paquetería de java	23
4.	Capitulo 4	25
	4.1. Instalación de la paquetería de java	
_		
		27
	5.1 Instalación de la paquetería de java	27





capítulo 1

Capitulo 1

1.1. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.





Administrar Pagos Admisión 2017-2018



Figura 1.1: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.2. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.

Administrar Pagos Admisión 2017-2018







Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.3. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.3: Administrar Pagos Admisión





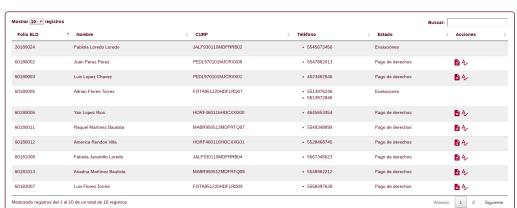
Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.4. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.4: Administrar Pagos Admisión





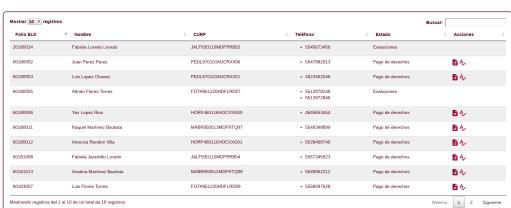
Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.5. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.5: Administrar Pagos Admisión





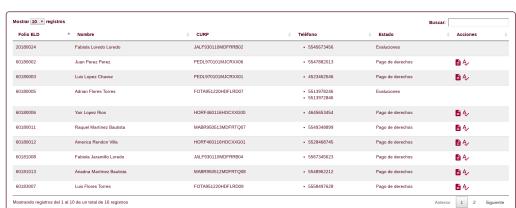
Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.6. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.6: Administrar Pagos Admisión





Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.7. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.7: Administrar Pagos Admisión





Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.8. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.8: Administrar Pagos Admisión





Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.9. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.9: Administrar Pagos Admisión





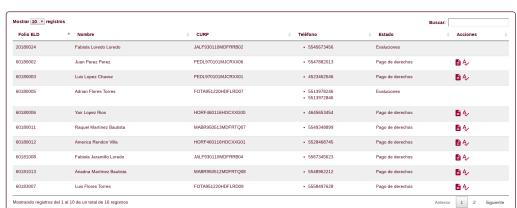
Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		

1.10. Introduccion

Para que la ELD pueda mantener actualizado el status de pago de los aspirantes que realizaron la operación en sucursal bancaria, es necesario que el **Contador General** actualice de forma manual dichos pagos, esto se logra con ayuda del archivo brindado por BANAMEX, el cual servirá para ser cargado en el sistema y este lo interprete y determine si es correcto, en caso de ser correcto actualizar el estado de pago de los aspirantes que allí se encuentran, de lo contrario mostrar las inconsistencias.

Una de las operaciones que brinda el sistema, es la actualización manual del status de pago de un aspirante, esta acción le permite tener el control para la actualización del status de pago de los aspirantes, esto con la finalidad de no depender del archivo de BANAMEX en caso de que presentaran conflictos con el mismo.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 1.10 Administrar Pagos Admisión.



Administrar Pagos Admisión 2017-2018

Figura 1.10: Administrar Pagos Admisión





Error	Posibles Soluciones		
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente. 		
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.		





capítulo 2

Capitulo 2

2.1. Marco Teórico

2.1.1. Big Data

El término Big Data se refiere a cantidades enormes de información, por ejemplo, la cantidad de información que se produce todos los días con el uso de una red social como Facebook, o la cantidad de datos que producen computadoras y dispositivos electrónicos que se auto monitorean mediante sensores. Esos volúmenes masivos de datos pueden ser utilizados para extraer conocimiento de ellos, y posteriormente atacar problemas que no sería posible resolver sin el Big Data.

Las 3 Vs del Big Data

Al ser el Big Data un concepto emergente y relativamente nuevo, se han tenido ciertas dificultades para definirlo de manera uniforme. Debido a las dimensiones de todo lo que conlleva entender Big Data, resultó conveniente entre los estudiosos del tema definir y acentuar las magnitudes que lo definen. Estas magnitudes se conocen como las 3 Vs del Big Data.

- 1. **Volumen:** Con Big Data, se tendrán que procesar enormes volúmenes de información. Este punto es importante ya que el crecimiento de la nececidad de almacenamiento de datos en el mundo moderno, crece de forma exponencial.
- 2. **Velocidad:** Usando Big Data, se abre la posibilidad de acceso y flujo de datos a velocidades que no se podrían conseguir de manera convencional.
- 3. Variedad: Procesamiento de datos de naturaleza heterogénea, es decir, múltiples tipos de datos.

Casos de uso del Big Data

- **Desarrollo de productos:** Compañías como Netflix y Procter & Gamble utilizan el Big Data para anticiparse a la demanda de los consumidores de sus productos. Utilizan modelos predictivos para sus nuevos productos clasificando atributos clave de sus anteriores productos modelando las relaciones entre esos atributos.
- Mantenimiento predictivo: Se pueden predecir fallas mecánicas en maquinaria que de otra forma quedarían fácilmente ignoradas. Mejorando así ampliamente la calidad y el costo del mantenimiento de dichos equipos.





- Experiencia de usuario: El uso de Big Data permite recopilar toda la información sobre el usuario y utilizarla a favor de su experiencia en un sitio o aplicación. Por ejemplo, sus búsquedas frecuentes, los sitios que visita, etc. Para de esta manera empezar a hacerle ofertas o anuncios personalizados, según sus intereses particulares.
- Machine Learning: Actualmente el machine learning es un área de mucho auge, ya que permite entrenar a las computadoras mediante conjuntos de datos de entrenamiento en lugar de programarlas. El Big Data facilita esa tarea.

2.1.2. Minería de datos

La minería de datos es el proceso de generar conocimiento a partir de un conjunto de información de gran tamaño. Para generar ese conocimiento se utilizan diferentes técnicas de análisis que detectan patrones y tendencias en la información que se está procesando. Si se intentara utilizar algún método de análisis tradicional, sería muy complicado o incluso imposible a veces encontrar patrones y tendencias útiles, ya que es muy probable que dentro de los datos existan relaciones muy complejas o simplemente la cantidad de datos sea demasiado grande.

La minería de datos puede utilizarse en escenarios como los que se enuncian a continuación:

- **Pronóstico:** Predicción de datos y eventos que vendrán a futuro a partir del comportamiento de conjuntos de datos que se tienen en el presente. Por ejemplo, predicción de ventas y tendencias de compra.
- Riesgo y probabilidad: Es un escenario muy común dentro de los negocios de Bussiness Intelligence. Por ejemplo, se llega a utilizar para encontrar puntos de equilibrio probable para escenarios de riesgo. Elección de los mejores clientes para la distribución de correo directo, determinación del punto de equilibrio probable para los escenarios de riesgo, y asignación de probabilidades a diagnósticos y otros resultados.
- **Recomendaciones:** Muy utilizado en sistemas como MercadoLibre o Amazon, en módulos que toman la información de búsqueda de cada usuario, la procesan y le arrojan recomendaciones de productos o servicios.
- **Búsqueda de secuencias:** Al igual que el escenario de **recomendaciones**, se utiliza mucho en sistemas de venta de artículos por internet. Se analiza el orden de los artículos que se meten a un carrito de compra para poder hacer predicciones útiles y generar conocimiento.
- **Agrupación:** Clasificación de los elementos de un conjunto de información para el posterior análisis de sus afinidades.

2.1.3. Árboles de decisión

Un árbol de decisión es un modelo de predicción que apoya al proceso de toma de decisiones. Esta herramienta tiene un campo de aplicación extremadamente amplio, pudiendo ir desde el área de finanzas hasta el área del aprendizaje de máquina. A partir de un conjunto de datos de entrada, se construyen los caminos, dentro del árbol, que llevarán a cada una de las decisiones posibles.

Los árboles de decisión están formados por los siguientes elementos:

- 1. **Nodos:** Un nodo es un punto del proceso en el que de acuerdo a ciertas condiciones o decisiones se redefine el rumbo del camino. Existen dos tipos de nodo:
 - 1.1. **Nodo de decisión:** Es un nodo en el cual se toma una decisión consciente de acuerdo a las necesidades del problema en cuestión. Estos nodos se representan con un cuadrado.
 - 1.2. **Nodo de incertidumbre:** Es un nodo en el que actúan las probabilidades y la heurística para definir el rumbo que tomará el camino que se hará dentro del árbol. Estos nodos se representan con un círculo
- 2. **Ramas:** Una rama es una de las respuestas o acciones que se toman a partir de la pregunta o escenario que se presentó en un nodo del cual salió esa rama. Una rama es el camino a otro nodo o escenario resultado de la decisión o evento que definió la rama. Este elemento se representa con una línea.
- 3. **Hojas:** Son escenarios finales, ya clasificados. No tienen ramificaciones, y son el resultado final de seguir un camino de decisiones, acciones y probabilidades. Este elemento se representa con un triángulo.





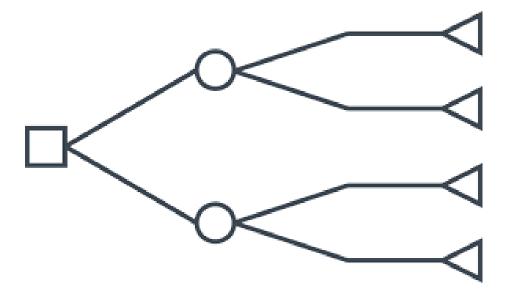


Figura 2.1: Representación general de un árbol de decisión

2.1.4. ID3 (Iterative Dichotomiser 3)

El algoritmo ID3 es uno de los algoritmos que utilizan árboles de decisión más populares. ID3 genera un árbol a partir de un conjunto de datos llamado **tabla de inducción**. Es útil para hacer toma de decisiones en escenarios binarios, es decir, que tienen 2 posibilidades finales.

El resultado de la ejecución de este algoritmo puede expresarse como un conjunto de reglas si-entonces.

Entropía

La entropía es la medida de la aleatoriedad. En otras palabras, la medidad e la impredictibilidad. La entropía es más alta cuando todos los eventos posibles en un escenario son igualmente probables, por ejemplo, al tirar una moneda al aire, se tiene 50 % de probabilidad de que caiga cara y 50 % de probabilidad de que caiga cruz, por lo que la entropía es de 1. Este parámetro comienca a decrecer cuando hay una probabilidad o probabilidades que parezcan aplastantes sobre las otras. Las fórmulas para calcular la entropía y la ganancia son las siguiente:

$$\operatorname{Entropia}(S) = \sum_{i=1}^n -p_i \log_2 p_i$$

Figura 2.2: Fórmula para calcular la entropía.

$$\operatorname{Gan} \operatorname{Inf}(S,A) = \operatorname{Entropia}(S) - \sum_{v \in V(A)} \frac{|Sv|}{|S|} \operatorname{Entropia}(Sv)$$

Figura 2.3: Fórmula para calcular la ganancia.





Ejecución del algoritmo ID3

Para ilustrar el funcionamiento del algoritmo ID3, utilizaremos la siguiente tabla de inducción que contiene información sobre factores que influyen al tomar la decisión de si un partido de tenis debería o no debería jugarse:

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	Caluroso	Alta	Ligero	No
2	Soleado	Caluroso	Alta	Fuerte	No
3	Nublado	Caluroso	Alta	Ligero	Sí
4	Lluvioso	Templado	Alta	Ligero	Sí
5	Lluvioso	Fresco	Normal	Ligero	Sí
6	Lluvioso	Fresco	Normal	Fuerte	No
7	Nublado	Fresco	Normal	Fuerte	Sí
8	Soleado	Templado	Alta	Ligero	No
9	Soleado	Fresco	Normal	Ligero	Sí
10	Lluvioso	Templado	Normal	Ligero	Sí
11	Soleado	Templado	Normal	Fuerte	Sí
12	Nublado	Templado	Alta	Fuerte	Sí
13	Nublado	Caluroso	Normal	Ligero	Sí
14	Lluvioso	Templado	Alta	Fuerte	No

Cuadro 2.1: Tabla de inducción para juegos de tenis.

Necesitaremos de la 2.2 fórmula para calcular la entropía y de la 2.3 fórmula para calcular la ganancia para poder proceder con la ejecución del algoritmo.

- Primeramente se calcula la entropía. La columna de **Decisión** consta de 14 instancias e incluye dos posibles valores: **sí** y **no**. Hay 9 decisiones con la etiqueta **sí** y 5 con la etiqueta **no**. Utilizando la fórmula de la entropía, se encuentra que el resultado es de 0.940.
- Después de los 5 factores disponibles, se busca el factor más dominante para tomar la decisión. Utilizando la fórmula de la ganancia tomando como parámetros la entropía de la columna **Decisión** y **Viento**.
- El atributo de viento tiene dos posibles valores: **fuerte** y **ligero**. Y al reflejar estos dos posibles valores en la fórmula, ésta se vería más o menos así: *Ganancia(Decisión, Viento) = Entropía(Decisión) [p(Decisión, Viento = Ligero)*Entropía(Decisión, Viento = Ligero)]-[p(Decisión, Viento = Fuerte)].*
 - Ahora se calcula (Decisión, Viento = Ligero) y (Decisión, Viento = Fuerte) respectivamente.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	Caluroso	Alta	Ligero	No
3	Nublado	Caluroso	Alta	Ligero	Sí
4	Lluvioso	Templado	Alta	Ligero	Sí
5	Lluvioso	Fresco	Normal	Ligero	Sí
8	Soleado	Templado	Alta	Ligero	No
9	Soleado	Fresco	Normal	Ligero	Sí
10	Lluvioso	Templado	Normal	Ligero	Sí
13	Nublado	Caluroso	Normal	Ligero	Sí

Cuadro 2.2: Tabla de para el factor de decisión Viento Ligero.

Dentro de la tabla de viento ligero hay 8 instancias en total, de las cuales 2 tienen la decisión final **no** y 6 tienen la decisión final **sí**. Al introducir estos datos en la 2.2 fórmula de la entropía, y obtenemos una entropía de 0.811.





Día	Aspecto	Temperatura	Humedad	Viento	Decisión
2	Soleado	Caluroso	Alta	Fuerte	No
6	Lluvioso	Fresco	Normal	Fuerte	No
7	Nublado	Fresco	Normal	Fuerte	Sí
11	Soleado	Templado	Normal	Fuerte	Sí
12	Nublado	Templado	Alta	Fuerte	Sí
14	Lluvioso	Templado	Alta	Fuerte	No

Cuadro 2.3: Tabla de para el factor de decisión Viento Fuerte.

En la tabla de viento fuerte encontramos 6 instancias divididas en 2 partes iguales: 3 tienen la decisión final **sí** y 3 tienen la decisión final **no**. Al sustituir estos datos en la 2.2 fórmula de la entropía, obtenemos una entropía de 1.

- Ahora que tenemos esos dos valores, podemos volver a la ecuación de la ganancia. El resultado queda expresado como Ganancia(Decisión, Viento) = 0.940-[(8/14)*0.811]-[(6/14)*1] = 0.048.
- En este punto se ha concluido ya el cálculo para el factor de decisión *Viento*. Ahora se necesita hacer lo mismo para todas las demas colúmnas/factores.
 - A modo de resumen:
 - 1. Ganancia (Decisión, Aspecto) = 0.246
 - 2. Ganancia (Decisión, Temperatura) = 0.029
 - 3. Ganancia(Decisión, Humedad) = 0.151
- Como podemos ver, el factor de decisión *Aspecto* es el que produce el valor de ganancia más alto. Por tal motivo, ese factor de decisión será el nodo raíz del árbol de decisión.
 - Ahora debemos probar todos los valores posibles que tiene el factor de decisión *Aspecto*:
 - Aspecto Nublado

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
3	Nublado	Caluroso	Alta	Ligero	Sí
7	Nublado	Fresco	Normal	Fuerte	Sí
12	Nublado	Templado	Alta	Fuerte	Sí
13	Nublado	Caluroso	Normal	Ligero	Sí

Cuadro 2.4: Tabla de días con aspecto nublado.

Podemos observar que siempre que el aspecto del día sea nublado, la decisión final sería sí.

Aspecto Soleado





Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	Caluroso	Alta	Ligero	No
2	Soleado	Caluroso	Alta	Fuerte	No
8	Soleado	Templado	Alta	Ligero	No
9	Soleado	Fresco	Normal	Ligero	Sí
11	Soleado	Templado	Normal	Fuerte	Sí

Cuadro 2.5: Tabla días con aspecto soleado.

En esta tabla tenemos 5 instancias para el aspecto **soleado**. Hay 3 de esas instancias que tienen como decisión final **sí** y 2 que tienen **no**. Por lo que los valores de ganancia del factor de decisión **Aspecto Soleado** con respecto a todos los demás factores de decisión quedarán de la siguiente manera:

- 1. Ganancia(Aspecto = Soleado, Temperatura) = 0.570
- 2. Ganancia(Aspecto = Soleado, Humedad) = 0.970
- 3. Ganancia(Aspecto = Soleado, Viento) = 0.019

En este punto, la humedad es el factor de decisión con mayor ganancia cuando el aspecto del día es soleado. Por lo que debemos probar todos los valores posibles para el factor de decisión *humedad*.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	Caluroso	Alta	Ligero	No
2	Soleado	Caluroso	Alta	Fuerte	No
8	Soleado	Templado	Alta	Ligero	No

Cuadro 2.6: Tabla días con aspecto soleado y humedad alta.

La decisión siempre será no cuando la humedad sea alta.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
9	Soleado	Fresco	Normal	Ligero	Sí
11	Soleado	Templado	Normal	Fuerte	Sí

Cuadro 2.7: Tabla días con aspecto soleado y humedad normal.

Por otro lado, la decisión siempre será sí cuando la humedad es normal.

De lo anterior concluimos que deberemos verificar la humedad y decidir si el aspecto del día es soleado.

Aspecto Lluvioso

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
4	Lluvioso	Templado	Alta	Ligero	Sí
5	Lluvioso	Fresco	Normal	Ligero	Sí
6	Lluvioso	Fresco	Normal	Fuerte	No
10	Lluvioso	Templado	Normal	Ligero	Sí
14	Lluvioso	Templado	Alta	Fuerte	No

Cuadro 2.8: Tabla de días con aspecto *lluvioso*.





Al evaluar los valores de ganancia de los días con aspecto **lluvioso** con respecto a los demás factores de decisión, se encuentra que el factor que genera una mayor ganancia es el viento. Por lo cual se tienen que checar todos los posibles valores de ese factor de decisión.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
4	Lluvioso	Templado	Alta	Ligero	Sí
5	Lluvioso	Fresco	Normal	Ligero	Sí
10	Lluvioso	Templado	Normal	Ligero	Sí

Cuadro 2.9: Tabla de días con aspecto *lluvioso* y con viento *ligero*.

De la tabla de aspecto **lluvioso** y viento **ligero** podemos deducir que siempre que existan estas dos condiciones al mismo tiempo, la decisión final será **sí**.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
6	Lluvioso	Fresco	Normal	Fuerte	No
14	Lluvioso	Templado	Alta	Fuerte	No

Cuadro 2.10: Tabla de días con aspecto *lluvioso*.

De la tabla de aspecto **lluvioso** y viento **fuerte** podemos deducir que siempre que existan estas dos condiciones al mismo tiempo, la decisión final será **no**.

• Finalmente la construcción de este árbol de decisión es la siguiente:

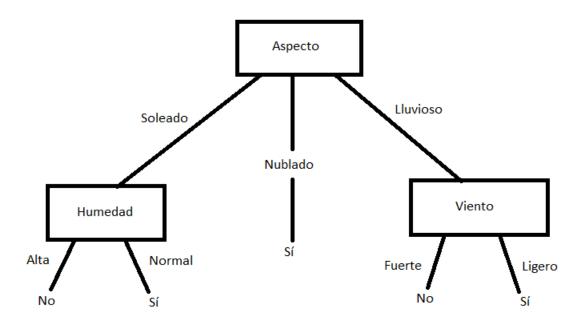


Figura 2.4: Árbol de decisión resultante.





2.1.5. C4.5

2.1.6. Spark

Es una herramienta de código abierto. Es un motor de análisis de datos unificado para Big Data y Machine Learning.

2.1.7. Hadoop

Es un framework de software que soporta aplicaciones distribuidas bajo una licencia libre. Permite a las aplicaciones trabajar con miles de nodos y petabytes de datos. Hadoop se inspiró en los documentos Google para MapReduce y Google File System (GFS). Hadoop utiliza su propio sistema de archivos HDFS, que divide archivos grandes y los distribuye en diferentes nodos para su procesamiento.

- 1. Solicite administrar los pagos admisión seleccionando la opción **Administración de pagos** del menú **??** y posteriormente la opción **Pagos Admisión** del menú **??**.
- 2. Se mostrará la pantalla 2.5 Administrar Pagos Admisión.



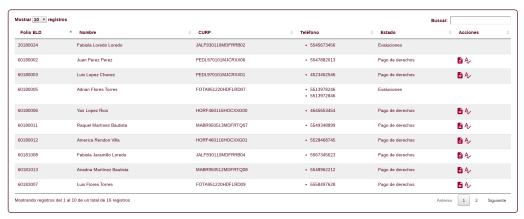


Figura 2.5: Administrar Pagos Admisión





Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	 Verifique que exista una convocatoria Publicada. Verifique que exista un periodo de pagos. Verifique que exista un periodo de pre-registro CENEVAL vigente.
El sistema muestra un mensaje indicando que no se ha realizado la asociación de fechas de CENEVAL y Psicométrico.	• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.





CAPÍTULO 3

Capitulo 3

3.1. Instalación de la paquetería de java

Primeramente será necesario instalar el open-jdk de java en el sistema, esto con el objetivo de permitir la ejecución de los algoritmos de minería de datos que serán ejecutados por el map reduce haciendo uso de java. Para realizar esta tarea realizaremos lo siguiente: Acceder a la terminal GNU con privilegios de root e ingresar el siguiente comando.

1.

root@maestro:~# apt-get install openjdk-8-jdk openjdk-8-jre

```
libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk-headless openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Paquetes sugeridos:
    default-jre libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo openjdk-8-source visualvm icedtea-8-plugin fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei fonts-indic

Se instalarán los siguientes paquetes NUEVOS:
    ca-certificates-java fonts-dejavu-extra java-common libgif7 libice-dev libxthread-stubs0-dev libxm-dev libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Se actualizarán los siguientes paquetes:
    libx11-6

1 actualizados, 22 nuevos se instalarán, 0 para eliminar y 573 no actualizados. Se necesita descargar 40.8 MB/41.4 MB de archivos.
Se utilizarán 165 MB de espacio de disco adicional después de esta operación. ¿Desea continuar? [S/n] s
Des:1 http://mx.archive.ubuntu.com/ubuntu xenial/main amd64 java-common all 0.56 ubuntu2 [7 742 B]
Des:2 http://mx.archive.ubuntu.com/ubuntu xenial-updates/main amd64 openjdk-8-jr e-headless amd64 8u181-b13-0ubuntu0.16.04.1 [27.0 MB]

46% [2 openjdk-8-jre-headless 22.2 MB/27.0 MB 82%]

703 kB/s 26s
```

Figura 3.1: Instalación de java en el nodo maestro

Este programa ocupara 165MB de espacio de disco y nos pedirá que confirmemos su instalación a lo que se contestará

2. 'S' para que proceda con la instalación. Una vez que esta termine se podrá consultar la versión de java en el sistema con el comando





root@maestro:~# java -version

```
root@maestro:/home/maestro# java -version
openjdk version "1.8.0_181"
OpenJDK Runtime Environment (build 1.8.0_181-8u181-b13-0ubuntu0.16.04.1-b13)
OpenJDK 64-Bit Server VM (build 25.181-b13, mixed mode)
```

Figura 3.2: Versión de java

capítulo 4

Capitulo 4

4.1. Instalación de la paquetería de java

Primeramente será necesario instalar el open-jdk de java en el sistema, esto con el objetivo de permitir la ejecución de los algoritmos de minería de datos que serán ejecutados por el map reduce haciendo uso de java. Para realizar esta tarea realizaremos lo siguiente: Acceder a la terminal GNU con privilegios de root e ingresar el siguiente comando.

1.

root@maestro:~# apt-get install openjdk-8-jdk openjdk-8-jre

```
libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk-headless openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Paquetes sugeridos:
    default-jre libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo openjdk-8-source visualvm icedtea-8-plugin fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei fonts-indic

Se instalarán los siguientes paquetes NUEVOS:
    ca-certificates-java fonts-dejavu-extra java-common libgif7 libice-dev libxthread-stubs0-dev libxm-dev libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Se actualizarán los siguientes paquetes:
    libx11-6

1 actualizados, 22 nuevos se instalarán, 0 para eliminar y 573 no actualizados. Se necesita descargar 40.8 MB/41.4 MB de archivos.

Se utilizarán 165 MB de espacio de disco adicional después de esta operación. ¿Desea continuar? [S/n] s

Des:1 http://mx.archive.ubuntu.com/ubuntu xenial/main amd64 java-common all 0.56 ubuntu2 [7 742 B]

Des:2 http://mx.archive.ubuntu.com/ubuntu xenial-updates/main amd64 openjdk-8-jr e-headless amd64 8u181-b13-0ubuntu0.16.04.1 [27.0 MB]

46% [2 openjdk-8-jre-headless 22.2 MB/27.0 MB 82%]

703 kB/s 26s
```

Figura 4.1: Instalación de java en el nodo maestro

Este programa ocupara 165MB de espacio de disco y nos pedirá que confirmemos su instalación a lo que se contestará

2. 'S' para que proceda con la instalación. Una vez que esta termine se podrá consultar la versión de java en el sistema con el comando





root@maestro:~# java -version

```
root@maestro:/home/maestro# java -version
openjdk version "1.8.0_181"
OpenJDK Runtime Environment (build 1.8.0_181-8u181-b13-0ubuntu0.16.04.1-b13)
OpenJDK 64-Bit Server VM (bu<u>i</u>ld 25.181-b13, mixed mode)
```

Figura 4.2: Versión de java

CAPÍTULO 5

Capitulo 5

5.1. Instalación de la paquetería de java

Primeramente será necesario instalar el open-jdk de java en el sistema, esto con el objetivo de permitir la ejecución de los algoritmos de minería de datos que serán ejecutados por el map reduce haciendo uso de java. Para realizar esta tarea realizaremos lo siguiente: Acceder a la terminal GNU con privilegios de root e ingresar el siguiente comando.

1.

root@maestro:~# apt-get install openjdk-8-jdk openjdk-8-jre

```
libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk-headless openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Paquetes sugeridos:
    default-jre libice-doc libsm-doc libxcb-doc libxt-doc openjdk-8-demo openjdk-8-source visualvm icedtea-8-plugin fonts-ipafont-gothic fonts-ipafont-mincho fonts-wqy-microhei fonts-wqy-zenhei fonts-indic

Se instalarán los siguientes paquetes NUEVOS:
    ca-certificates-java fonts-dejavu-extra java-common libgif7 libice-dev libxthread-stubs0-dev libxm-dev libx11-dev libx11-doc libxau-dev libxcb1-dev libxdmcp-dev libxt-dev openjdk-8-jdk openjdk-8-jdk-headless openjdk-8-jre openjdk-8-jre-headless x11proto-core-dev x11proto-input-dev x11proto-kb-dev xorg-sgml-doctools xtrans-dev

Se actualizarán los siguientes paquetes:
    libx11-6

1 actualizados, 22 nuevos se instalarán, 0 para eliminar y 573 no actualizados. Se necesita descargar 40.8 MB/41.4 MB de archivos.

Se utilizarán 165 MB de espacio de disco adicional después de esta operación. ¿Desea continuar? [S/n] s

Des:1 http://mx.archive.ubuntu.com/ubuntu xenial/main amd64 java-common all 0.56 ubuntu2 [7 742 B]

Des:2 http://mx.archive.ubuntu.com/ubuntu xenial-updates/main amd64 openjdk-8-jr e-headless amd64 8u181-b13-0ubuntu0.16.04.1 [27.0 MB]

46% [2 openjdk-8-jre-headless 22.2 MB/27.0 MB 82%]

703 kB/s 26s
```

Figura 5.1: Instalación de java en el nodo maestro

Este programa ocupara 165MB de espacio de disco y nos pedirá que confirmemos su instalación a lo que se contestará

2. 'S' para que proceda con la instalación. Una vez que esta termine se podrá consultar la versión de java en el sistema con el comando





root@maestro:~# java -version

```
root@maestro:/home/maestro# java -version
openjdk version "1.8.0_181"
OpenJDK Runtime Environment (build 1.8.0_181-8u181-b13-0ubuntu0.16.04.1-b13)
OpenJDK 64-Bit Server VM (bu<u>i</u>ld 25.181-b13, mixed mode)
```

Figura 5.2: Versión de java