

Luminus

—

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

17 de octubre de 2018

1. Capítulo 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	6
1.7. Introduccion	7
1.8. Introduccion	8
1.9. Introduccion	9
1.10. Introduccion	10
2. Capítulo 2	13
2.1. Marco Teórico	13
2.1.1. Big Data	13
2.1.2. Minería de datos	14
2.1.3. Árboles de decisión	14
2.1.4. ID3 (Iterative Dichotomiser 3)	15
2.1.5. C4.5	20
2.1.6. Algoritmo KNN (K-Nearest Neighbor)	25
2.1.7. Algoritmo K-Means	26
2.1.8. MapReduce	28
2.1.9. Hadoop	28
2.1.10. Apache Spark	30
3. Capítulo 3	35
3.1. Instalación de la paquetería de java	35
4. Capítulo 4	37
4.1. Instalación de la paquetería de java	37
5. Capítulo 5	39
5.1. Instalación de la paquetería de java	39

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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

12

Siguiente

Carros Archivos

[Cargar Archivo](#)

Figura 1.1: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

12

Siguiente

Cargar Archivo

Figura 1.3: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

1

2

Siguiente

Cargar Archivo

Figura 1.4: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loreda Loreda	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loreda	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

1

2

Siguiente

Cargar Archivo

Figura 1.5: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loreda Loreda	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loreda	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

12

Siguiente

Cargar Archivo

Figura 1.6: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

12

Siguiente

Cargar Archivo

Figura 1.7: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

1

2

Siguiente

Cargar Archivo

Figura 1.8: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loreda Loreda	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loreda	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

1

2

Siguiente

Cargar Archivo

Figura 1.9: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • 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

Mostrar 10 registros

Buscar:

Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones	
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos	 
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos	 
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones	
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos	 
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos	 
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos	 
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos	 
60181013	Ariadna Martinez Bautista	MABR950512MDFRTQ08	• 5548962212	Pago de derechos	 
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos	 

Mostrando registros del 1 al 10 de un total de 16 registros

Anterior

12

Siguiente

Cargar Archivo

Figura 1.10: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none">• 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.	<ul style="list-style-type: none">• Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.

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 necesidad 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 Business 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.

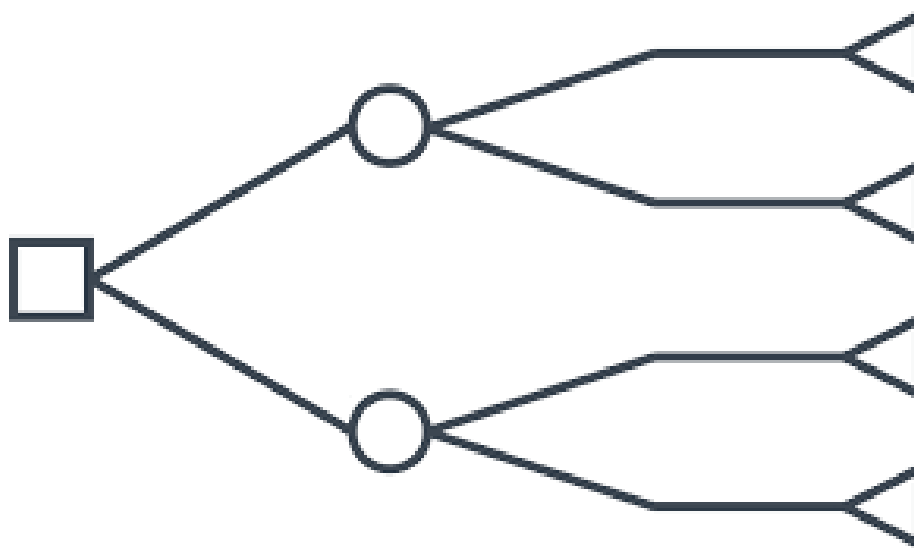


Figura 2.1: Representación general de un árbol de decisión

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 medida de la impredecibilidad. 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 comienza 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 siguientes:

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

Figura 2.2: Fórmula para calcular la entropía.

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

Figura 2.3: Fórmula para calcular la ganancia.

Ejemplo de 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 ID3.

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: $Ganancia(Decisión, Viento) = Entropía(Decisión) - [p(Decisión, Viento = Ligero) * Entropía(Decisión, Viento = Ligero)] - [p(Decisión, Viento = Fuerte) * Entropía(Decisión, Viento = Fuerte)]$.

- 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 columnas/factores.

- A modo de resumen:

- Ganancia(Decisión, Aspecto) = 0.246
- Ganancia(Decisión, Temperatura) = 0.029
- 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:

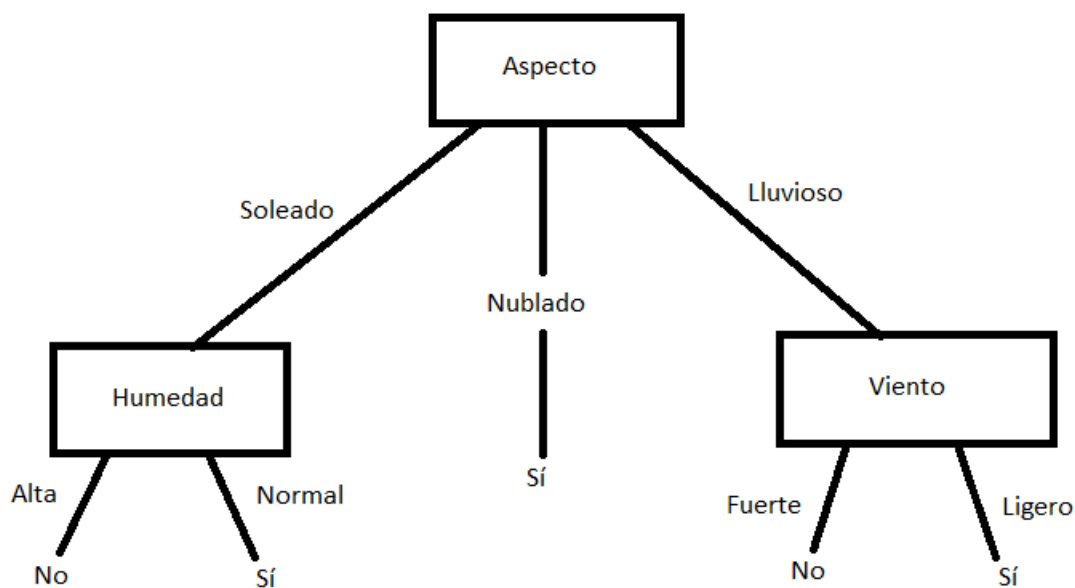


Figura 2.4: Árbol de decisión resultante.

2.1.5. C4.5

El algoritmo C4.5 es la evolución del algoritmo ID3. Éste genera un árbol de decisión a partir de un conjunto de datos de entrada de manera recursiva, al igual que su precursor. Sin embargo, aunque ID3 y C4.5 son algoritmos muy semejantes, existen ciertas diferencias:

- C4.5 permite trabajar con valores continuos, mientras que ID3 solamente trabaja con valores discretos.
- Los árboles de C4.5 son más compactos, y esto se debe a que cada nodo hoja engloba un conjunto de clases y no una sola clase particular.
- C4.5 es más eficiente computacionalmente hablando.
- C4.5 utiliza un nuevo parámetro llamado **Gain Ratio**, en lugar de la ganancia simple. Y este a su vez requiere de la ganancia y de otro parámetro nuevo llamado **SplitInfo** cuyas fórmulas se enuncian a continuación:

$$GainRatio(A) = \frac{Gain(A)}{SplitInfo(A)}$$

Figura 2.5: Fórmula para obtener el Gain Ratio

$$SplitInfo_A(D) = - \sum_{j=1}^v \frac{|D_j|}{|D|} \times \log_2 \left(\frac{|D_j|}{|D|} \right)$$

Figura 2.6: Fórmula para obtener SplitInfo

Ejemplo de ejecución del algoritmo C4.5

Para mostrar la forma en la que se ejecuta el algoritmo C4.5 y los pasos que se deben llevar a cabo, se utilizará la siguiente tabla de inducción:

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	85	85	Ligero	No
2	Soleado	80	90	Fuerte	No
3	Nublado	83	78	Ligero	Sí
4	Lluvioso	70	96	Ligero	Sí
5	Lluvioso	68	80	Ligero	Sí
6	Lluvioso	65	70	Fuerte	No
7	Nublado	64	65	Fuerte	Sí
8	Soleado	72	95	Ligero	No
9	Soleado	69	70	Ligero	Sí
10	Lluvioso	75	80	Ligero	Sí
11	Soleado	75	70	Fuerte	Sí
12	Nublado	72	90	Fuerte	Sí
13	Nublado	81	75	Ligero	Sí
14	Lluvioso	80	80	Fuerte	No

Cuadro 2.11: Tabla de inducción para juegos de tenis C4.5.

- Al igual que con el ejemplo **Ejemplo de ejecución del algoritmo ID3**, lo primero que se hace es calcular la entropía general. Hay 15 instancias de las cuales 9 tienen una decisión final de **sí** y 5 tienen **no**. Al sustituir los valores correspondientes en la **2.2 fórmula para calcular la entropía** el resultado que obtenemos es **0.940**.

- En C4.5 se utilizan **Gain Ratios** (ratos de ganancia), mientras que en ID3 se utilizan ganancias.
- Empezaremos por analizar el atributo de **Viento**.

Se tienen 8 instancias de *viento ligero*, dos de ellas concluyen en un **no**, y las otras 6 concluyen en **sí** por lo que:

- Entropía(Decisión, Viento = Ligero) = 0.811

- Entropía(Decisión, Viento = Fuerte) = 1

- Ganancia(Decisión, Viento) = 0.049

Existen 6 instancias de viento fuerte por lo que:

- SplitInfo(Decisión, Viento) = 0.985

- GainRatio(Decisión, Viento) = Gain(Decisión, Viento) / SplitInfo(Decisión, Viento) = 0.049

- Continuamos analizando ahora el atributo **Aspecto**.

Se tienen 5 instancias para el aspecto *soleado*, de las cuales 3 concluyen en **no** y las otras 2 concluyen en **sí**. Calculando sus valores de entropías, ganancia, SplitInfo y GainRatio:

- Entropía(Decisión, Aspecto = Soleado) = 0.970

- Entropía(Decisión, Aspecto = Nublado) = 0

3. Entropía(Decisión, Aspecto = Lluvioso) = 0.970

4. Ganancia(Decisión, Aspecto) = 0.246

Hay 5 instancias para *soleado*, 4 instancias para *nublado* y 5 para *lluvioso*, por lo que:

5. SplitInfo(Decisión, Aspecto) = 1.577

6. GainRatio(Decisión, Aspecto) = 0.155

• Procedemos a analizar el atributo de humedad. Cuyo caso es diferente al de los demás atributos, ya que este es un atributo continuo. Necesitamos convertir valores continuos a valores nominales (como todos los demás atributos). C4.5 propone hacer una división binaria a partir de algún valor que podamos tomar como umbral. El umbral debe ser el valor que mayor ganancia ofrezca para ese atributo. Para esto, primero se deben ordenar las instancias de humedad de menor a mayor.

Día	Humedad	Decisión
7	65	Sí
6	70	No
9	70	Sí
11	70	Sí
13	75	Sí
3	78	Sí
5	80	Sí
10	80	Sí
14	80	No
1	85	No
2	90	No
12	90	Sí
8	95	No
4	96	Sí

Cuadro 2.12: Tabla de l atributo *humedad* ordenada de menor a mayor.

Ahora debemos recorrer todos los valores de humedad y separar el conjunto de datos en dos partes. Se calcularán la **Ganancia** y el **GainRatio** y el valor que maximice la ganancia será el umbral.

1. Se propone 65 como umbral.

1.1. Entropía(Decisión, Humedad \leq 65) = 0

1.2. Entropía(Decisión, Humedad $>$ 65) = 0.961

1.3. Ganancia(Decisión, Humedad $<>$ 65) = 0.048

1.4. SplitInfo(Decisión, Humedad $<>$ 65) = 0.371

1.5. GainRatio(Decisión, Humedad $<>$ 65) = 0.126

2. Se propone 70 como umbral.

2.1. Entropía(Decisión, Humedad \leq 70) = 0.811

2.2. Entropía(Decisión, Humedad $>$ 70) = 0.970

2.3. $Ganancia(Decisión, Humedad <>70) = 0.014$

2.4. $SplitInfo(Decisión, Humedad <>70) = 0.863$

2.5. $GainRatio(Decisión, Humedad <>70) = 0.016$

3. Se propone 75 como umbral.

3.1. $Entropía(Decisión, Humedad \leq 75) = 0.721$

3.2. $Entropía(Decisión, Humedad >75) = 0.991$

3.3. $Ganancia(Decisión, Humedad <>75) = 0.045$

3.4. $SplitInfo(Decisión, Humedad <>75) = 0.940$

3.5. $GainRatio(Decisión, Humedad <>75) = 0.047$

4. Se continúa haciendo lo mismo para cada valor nuevo de humedad que se vaya encontrando en la tabla. Resumiendo:

5. $Ganancia(Decisión, Humedad <>78) = 0.090$

6. $Ganancia(Decisión, Humedad <>80) = 0.107$

7. $Ganancia(Decisión, Humedad <>85) = 0.027$

8. $Ganancia(Decisión, Humedad <>90) = 0.016$

• Como podemos ver, el valor que maximiza la ganancia es el de 80. Lo que significa que ahora se debe comparar los otros valores nominales con el valor 80 del atributo *humedad* para crear una rama en nuestro árbol. De esta forma podemos resumir los resultados en la siguiente tabla:

Atributo	Ganancia	GainRatio
Viento	0.049	0.049
Aspecto	0.246	0.155
Humedad <>80	0.101	0.107

Cuadro 2.13: Comparación de las ganancias entre atributos.

• Al igual que en el ejemplo del algoritmo ID3, el atributo **aspecto** es el nodo raíz del árbol de decisión, por lo que ahora se deben analizar todos los posibles valores que puede adquirir dicho atributo.

• Aspecto Soleado.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
1	Soleado	85	85	Ligero	No
2	Soleado	80	90	Fuerte	No
8	Soleado	72	95	Ligero	No
9	Soleado	69	70	Ligero	Sí
11	Soleado	75	70	Fuerte	Sí

Cuadro 2.14: Tabla de inducción para el atributo *Aspecto* con el valor *Soleado*.

Ya hemos dividido la humedad a partir de su punto de umbral que es 80. Podemos observar en la siguiente tabla que la decisión final será **no**, si la humedad es mayor a 80 y el aspecto del día es soleado. La decisión final será **sí**, si

la humedad es menor o igual a 80 para un día soleado.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
3	Nublado	83	78	Ligero	Sí
7	Nublado	64	65	Fuerte	Sí
12	Nublado	72	90	Fuerte	Sí
13	Nublado	81	75	Ligero	Sí

Cuadro 2.15: Tabla de inducción para el atributo *Aspecto* con el valor *Nublado*.

Cuando el aspecto del día es **nublado**, no importa ninguna otra condición; ni temperatura, ni humedad. La decisión final siempre será **sí**.

Día	Aspecto	Temperatura	Humedad	Viento	Decisión
4	Lluvioso	70	96	Ligero	Sí
5	Lluvioso	68	80	Ligero	Sí
6	Lluvioso	65	70	Fuerte	No
10	Lluvioso	75	80	Ligero	Sí
14	Lluvioso	80	80	Fuerte	No

Cuadro 2.16: Tabla de inducción para el atributo *Aspecto* con el valor *Lluvioso*.

Cuando el aspecto es **lluvioso**, la decisión será **sí** cuando el viento es ligero, y será **no** cuando sea fuerte.

- De todo lo anterior se concluye un árbol de decisión con la siguiente estructura:

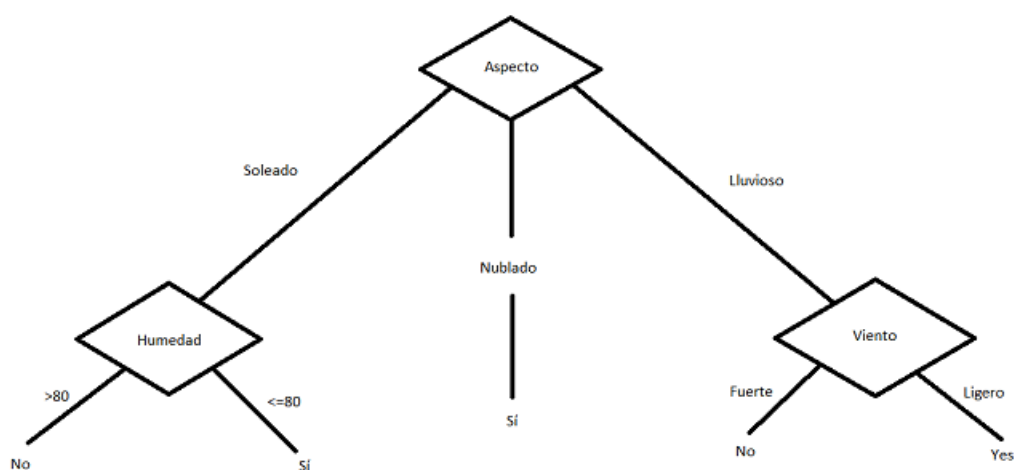


Figura 2.7: Árbol de decisión construido con C4.5.

2.1.6. Algoritmo KNN (K-Nearest Neighbor)

El algoritmo KNN es un algoritmo de aprendizaje no supervisado en el que se busca clasificar un punto en una categoría con la ayuda de un conjunto de entrenamiento.

Los pasos que sigue el algoritmo KNN se enumeran a continuación:

1. Se calcula la similitud entre puntos basándose en una función de distancia.
2. Se encuentran los K vecinos más cercanos.

Ejemplo de ejecución del algoritmo KNN

Con base en la siguiente información que relaciona el peso, la altura y la talla de playera de personas, se hará la ejecución del algoritmo KNN.

Altura (cm)	Peso (kg)	Talla
158	58	M
158	59	M
158	63	M
160	59	M
160	60	M
163	60	M
163	61	M
160	64	L
163	64	L
165	61	L
165	62	L
165	65	L
168	62	L
168	63	L
168	66	L
170	63	L
170	64	L
170	68	L

Cuadro 2.17: Conjunto de entrenamiento para el algoritmo KNN.

Nuevo dato: Altura = 161 cm, Peso = 61 kg

- Se calcula la distancia Euclidiana entre el dato de entrada y cada uno de los datos del conjunto de entrenamiento, utilizando la siguiente fórmula.

$$d_E(P, Q) = \sqrt{(p_1 - q_1)^2 + (p_2 - q_2)^2 + \dots + (p_n - q_n)^2} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}.$$

Figura 2.8: Fórmula para calcular la distancia Euclidiana.

Obtenemos lo siguiente:

Altura (cm)	Peso (kg)	Talla	Distancia
158	58	M	4.2
158	59	M	3.6
158	63	M	3.6
160	59	M	2.2
160	60	M	1.4
163	60	M	2.2
163	61	M	2.0
160	64	L	3.2
163	64	L	3.6
165	61	L	4.0
165	62	L	4.1
165	65	L	5.7
168	62	L	7.1
168	63	L	7.3
168	66	L	8.6
170	63	L	9.2
170	64	L	9.5
170	68	L	11.4

Cuadro 2.18: Tabla del conjunto de entrenamiento con la columna **Distancia** ya calculada.

- Se toma la distancia más pequeña para determinar al vecino más cercano de entre todas las distancias Euclidianas calculadas previamente. Esta distancia es **1.4**.
- Se define K como el número de vecinos más cercanos que queramos conocer. Para este caso definiremos K igual a 5. Por lo que se tomarán los 5 valores más pequeños.

Altura (cm)	Peso (kg)	Talla	Distancia
160	60	M	1.4
163	61	M	2.0
163	60	M	2.2
160	59	M	2.2
160	64	L	3.2

Cuadro 2.19: Instancias de la tabla con menor distancia al nuevo dato.

2.1.7. Algoritmo K-Means

El algoritmo K-Means es un algoritmo de agrupamiento que utiliza aprendizaje no supervisado, el cual es utilizado cuando se tienen datos no etiquetados, es decir sin categorías o grupos definidos. El objetivo de este algoritmo es encontrar grupos en los datos, con el número de grupos se representa la variable K . El algoritmo arrojará como resultado final lo siguiente:

1. Los centroides de los K grupos.
2. Etiquetas para los datos de entrenamiento. Cada dato es asignado a un solo grupo.

Ejemplo de ejecución del algoritmo K-Means

Para ilustrar el funcionamiento del algoritmo K-means, se utilizará el siguiente conjunto de entrenamiento correspondiente a la puntuación de 7 individuos en 2 pruebas:

Individuo	Prueba 1	Prueba 2
1	1.0	1.0
2	1.5	2.0
3	3.0	4.0
4	5.0	7.0
5	3.5	5.0
6	4.5	5.0
7	3.5	4.5

Cuadro 2.20: Conjunto de entrenamiento para el algoritmo K-Means.

- El conjunto de datos se agrupa en dos grupos distintos. Para esto, se toman los valores de la *Prueba 1* y *Prueba 2* de los individuos que tengan estos valores más lejanos, es decir:

Grupo	Individuo	Vector (Centroide)
Grupo 1	1	(1.0, 1.0)
Grupo 2	4	(5.0, 7.0)

Cuadro 2.21: Partición inicial.

- Los individuos restantes ahora son examinados secuencialmente y son ubicados en el grupo al que son más cercanos en términos de distancia Euclidiana. El vector es recalculado cada que un nuevo miembro es agregado.

Iteración	Individuo	Vector (Centroide)
1	1	(1.0, 1.0)
2	1, 2	(1.2, 1.5)
3	1, 2, 3	(1.8, 2.3)
4	1, 2, 3	(1.8, 2.3)
5	1, 2, 3	(1.8, 2.3)
6	1, 2, 3	(1.8, 2.3)

Cuadro 2.22: Tabla de individuos agregados secuencialmente al **Grupo 1**.

Iteración	Individuo	Vector (Centroide)
1	4	(5.0, 7.0)
2	4	(5.0, 7.0)
3	4	(5.0, 7.0)
4	4, 5	(4.2, 6.0)
5	4, 5, 6	(4.3, 5.7)
6	4, 5, 6, 7	(4.1, 5.4)

Cuadro 2.23: Tabla de individuos agregados secuencialmente al **Grupo 2**.

- Ahora la partición inicial ha cambiado, y los dos grupos tienen las siguientes características:

Grupo	Individuo	Vector (Centroide)
Grupo 1	1, 2, 3	(1.8, 2.3)
Grupo 2	4, 5, 6, 7	(4.1, 5.4)

Cuadro 2.24: Partición inicial modificada.

- Sin embargo, no podemos asegurarnos de que esas son las clasificaciones correctas para cada dato. Así que comparamos la distancia de cada individuo a su propio grupo y al grupo opuesto:

Individuo	Distancia al Grupo 1	Distancia al Grupo 2
1	1.5	5.4
2	0.4	4.3
3	2.1	1.8
4	5.7	1.8
5	3.2	0.7
6	3.8	0.6
7	2.8	1.1

Cuadro 2.25: Cálculo de distancia de cada dato a su grupo y al opuesto.

- Solamente el individuo 3 está más cerca al centroide del grupo opuesto (Grupo 2) que de su propio grupo (Grupo 1). Por lo que el individuo 3 se reubica en el Grupo 2. Resultando la siguiente partición:

Grupo	Individuo	Vector (Centroide)
Grupo 1	1, 2	(1.8, 1.5)
Grupo 2	3, 4, 5, 6, 7	(3.9, 5.1)

Cuadro 2.26: Partición final.

- Las iteraciones continuarían hasta que ya no haya más reubicaciones.
- Es probable que el algoritmo no encuentre una solución final.

2.1.8. MapReduce

MapReduce es un paradigma de programación que permite escalabilidad masiva a través de cientos o miles de servidores en un clúster Hadoop. *MapReduce* se refiere a dos tareas distintas y separadas. La primera, *Map*, consiste en realizar mapeos. Ésta convierte un conjunto de datos en otro conjunto de datos diferente en el que los elementos individuales son separados en tuplas. La segunda, *Reduce*, toma los datos que arroja de *Map*, y combina esas tuplas en un conjunto más pequeño de tuplas. MapReduce es el corazón de Hadoop.

2.1.9. 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.

Características principales de Hadoop

- **Procesamiento distribuido:** Hadoop distribuye los datos en los diferentes nodos que formen parte de la arquitectura de clúster que estén haciendo uso de él. Pretende paralelizar tareas de procesamiento de datos.
- **Eficiencia:** Mediante la paralelización, se consigue una ganancia considerable en el tiempo de procesamiento de información.
- **Económico:** Propicia un ambiente fácilmente escalable en el que resulta sencillo añadir nodos de manera horizontal conforme se vaya requiriendo.
- **Código abierto:** Es un proyecto de los llamados **open source**.
- **Tolerancia a fallos:** Utiliza replicación de datos haciendo uso de **HDFS (Hadoop Distributed File System)**. Si un nodo falla o cae, hay nodos de respaldo que permiten mantener el ambiente en funcionamiento.

Arquitectura de Hadoop

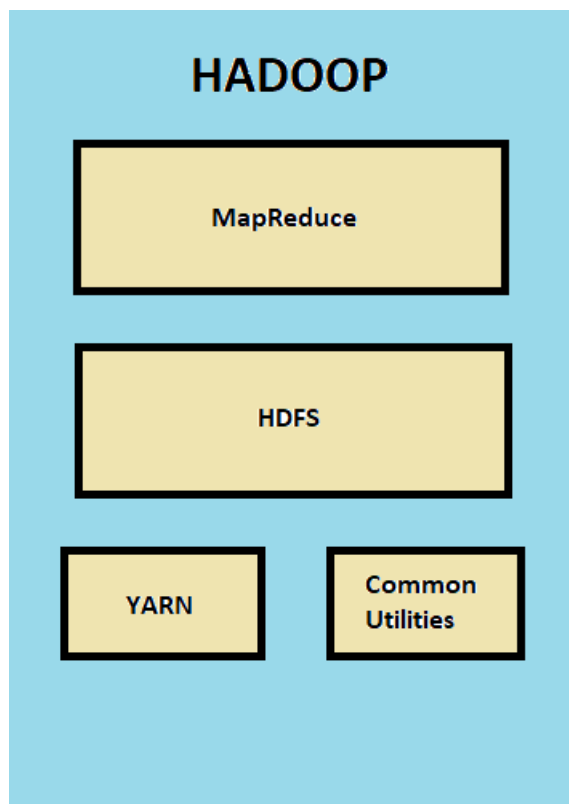


Figura 2.9: Arquitectura básica de Hadoop.

- **MapReduce:** Es el módulo que permite ejecutar cómputo distribuido. Pretende la paralelización de procesos.
- **HDFS (Hadoop Distributed File System:)** Es el sistema de archivos distribuido que utiliza Hadoop. Éste parte los datos en fragmentos y los almacena en los nodos que conforman el clúster donde Hadoop esté ejecutándose.
- **YARN:** Es el gestor de recursos de Hadoop.
- **Common Utilities:** Librerías y código necesario para ejecutar Hadoop.

Arquitectura de un clúster Hadoop

Habitualmente, un clúster Hadoop tiene la estructura **maestro - esclavo**. Lo que significa que hay un nodo **maestro** que estará coordinando la ejecución de tareas y las asignará a los nodos **esclavos**.

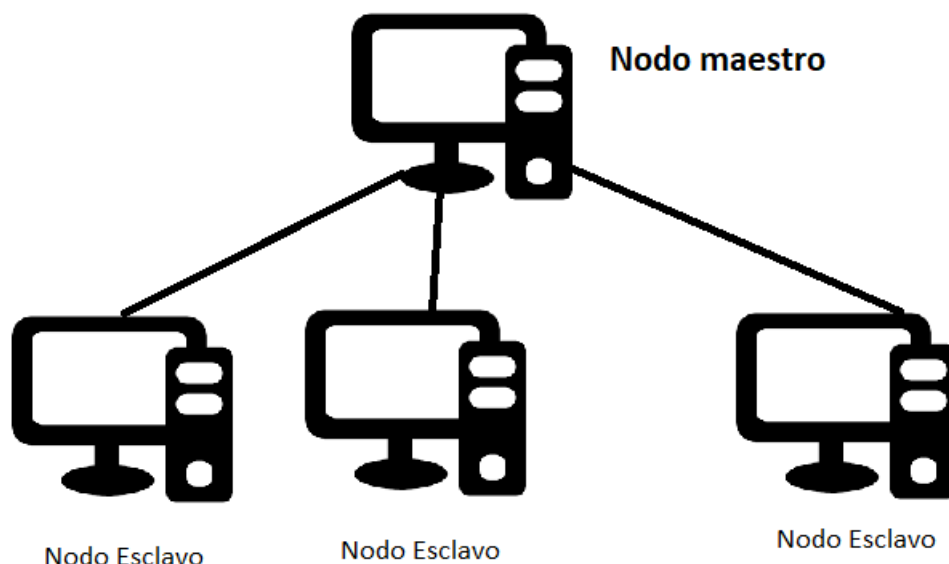


Figura 2.10: Arquitectura de maestros y esclavos.

Distribuciones Comerciales de Hadoop

Existen varias opciones actualmente en el mercado para hacer uso de las capacidades de Hadoop, entre las cuales destacan 3 principalmente:

- **Cloudera:** Fue la primera de todas las distribuciones comerciales de Hadoop. Cuenta con una herramienta llamada Cloudera Manager que permite gestionar el clúster.
- **Hortonworks:** Es la distribución de Hadoop más nueva. Permite instalar el clúster mediante un cliente de virtualización.
- **Microsoft Azure:** Es una distribución desarrollada por Microsoft que permite tener las máquinas del clúster en la nube.

2.1.10. Apache Spark

Es un motor de análisis de datos unificado para Big Data y Machine Learning. Que se basa en Hadoop Map Reduce. Es una herramienta de código abierto que permite **dividir y ejecutar tareas de manera paralela**. Esta cualidad de Spark de poder paralelizar el trabajo se debe a que éste software en la gran mayoría de los casos es ejecutado en sistemas con arquitectura de clúster.

Características principales de Spark

- Integrado con Apache Hadoop.

- Ofrece un desempeño veloz.
- El almacenamiento de datos se administra en memoria. Se reducen mucho los tiempos de ejecución ya que hay muchas menos operaciones de lectura y escritura en disco.
- Puede ejecutar algoritmos escritos en Java, Scala, Python y R.
- Permite procesamiento en tiempo real.

Ventajas de usar Apache Spark

Utilizar Spark para el procesamiento de volúmenes de información de gran tamaño ofrece varios beneficios. Los principales beneficios son los siguientes:

- **Velocidad:** Gran velocidad en el procesamiento de información. Esto se debe a lo ya mencionado anteriormente sobre la gestión de datos desde memoria.
- **Potencia:** Spark nos permite aprovechar el hardware de los equipos que lo estén utilizando.
- **Fácil uso:** A diferencia de Hadoop, que requería de un amplio conocimiento de MapReduce y de Java, Spark permite usar lenguajes de más alto nivel como Python y Scala además de Java.
- **Integración SQL:** Como resultado de contener el módulo Spark SQL, es posible realizar consultas en conjuntos de datos semi-estructurados utilizando lenguaje SQL.
- **Constante mejora del propio sistema:** Al ser un proyecto de código abierto, cada vez hay más personas que contribuyen a la mejora y ampliación de los alcances de Spark.
- **Escalabilidad:** Spark permite que incrementar el tamaño del clúster conforme se vaya necesitando.

Componentes de Apache Spark

Podría decirse que Spark es un conjunto de módulos que nos permiten generar conocimiento usando diferentes técnicas y tecnologías.

El diagrama mostrado a continuación ilustra los principales módulos o **componentes** que conforman Apache Spark.

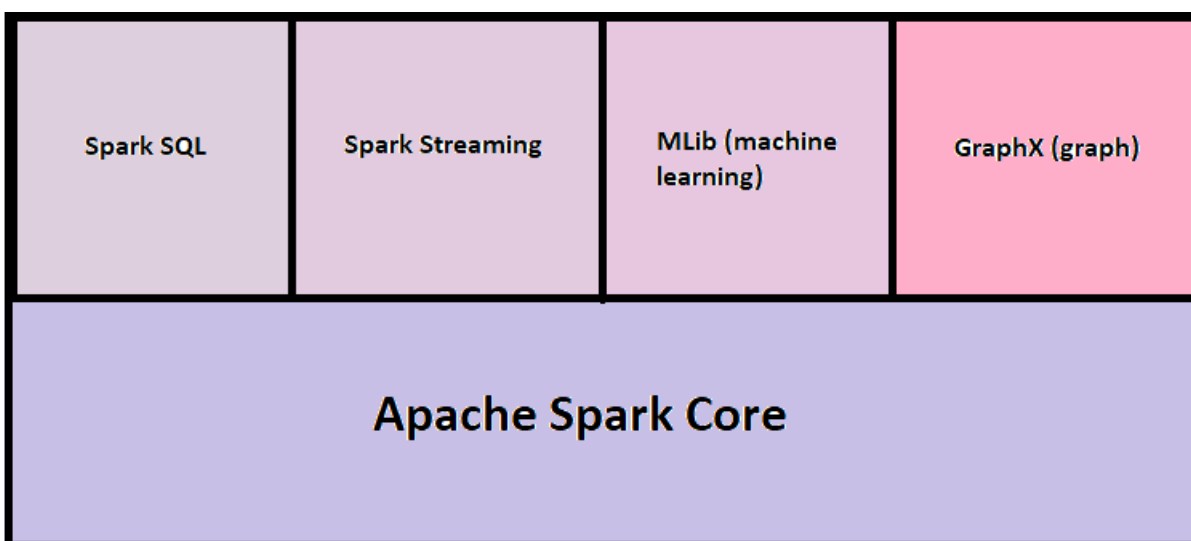


Figura 2.11: Componentes que conforman Apache Spark.

- **Spark Core:** Es el núcleo de Spark. Contiene librerías que se utilizan en todos los demás módulos.
 - **Spark SQL:** Módulo para el procesamiento de datos estructurados y semi-estructurados. Esto se conoce como *SchemaRDD*. Este módulo hace posible el uso de lenguaje SQL para hacer consultas.
 - **Spark Streaming:** Este módulo hace posible el procesamiento en tiempo real. Hace posible el flujo de gran cantidad de datos a alta velocidad.
 - **MLib:** Este módulo contiene herramientas y algoritmos muy variados para usar de manera fácil, práctica y escalable al *machine learning*.
 - **GraphX:** Este módulo permite el procesamiento gráfico. No pinta gráficos, sino que realiza operaciones con grafos.
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 [2.12 Administrar Pagos Admisión](#).

Administrar Pagos Admisión 2017-2018

Mostrar 10 registros						Buscar:
Folio ELD	Nombre	CURP	Teléfono	Estado	Acciones	
20180024	Fabiola Loredo Loredo	JALF930110MDFRRB02	• 5545673456	Evaluaciones		
60180002	Juan Perez Perez	PEDL970101MJCRXX06	• 5547882013	Pago de derechos		
60180003	Luis Lopez Chavez	PEDL970101MJCRXX01	• 4523462546	Pago de derechos		
60180005	Adrian Flores Torres	FOTA951220HDFLRD07	• 5513978246 • 5513972846	Evaluaciones		
60180006	Yair Lopez Rios	HORF460116HOCXXG00	• 4645653454	Pago de derechos		
60180011	Raquel Martinez Bautista	MABR950513MDFRTQ07	• 5549348899	Pago de derechos		
60180012	America Rendon Villa	HORF460116HOCXXG01	• 5528468745	Pago de derechos		
60181008	Fabiola Jaramillo Loredo	JALF930110MDFRRB04	• 5567345623	Pago de derechos		
60181013	Ariadna Martinez Bautista	MABR950513MDFRTQ08	• 5548962212	Pago de derechos		
60183007	Luis Flores Torres	FOTA951220HDFLRD09	• 5558497628	Pago de derechos		
Mostrando registros del 1 al 10 de un total de 16 registros						Anterior 1 2 Siguiente

[Cargar Archivo](#)

Figura 2.12: Administrar Pagos Admisión

Errores Comunes

Error	Posibles Soluciones
El sistema muestra un mensaje indicando que falta información para realizar la operación.	<ul style="list-style-type: none"> • 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.	<ul style="list-style-type: none"> • Verifique que la Coordinación de Control Escolar haya asociado las fechas CENEVAL y Psicométrico.

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
```

```

root@maestro: /home/maestro
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
libpthread-stubs0-dev libsm-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

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
```

```

root@maestro: /home/maestro
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
libpthread-stubs0-dev libsm-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 (build 25.181-b13, mixed mode)
```

Figura 4.2: Versión de java

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
```

```

root@maestro: /home/maestro
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
libpthread-stubs0-dev libsm-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 (build 25.181-b13, mixed mode)
```

Figura 5.2: Versión de java