

# Selección de actividades extraescolares aplicando minería de datos

***Claudia Rodríguez Lemus***

Instituto Tecnológico de Roque

*claulemus@itroque.edu.mx*

***Ma. Guadalupe López Bedolla***

Instituto Tecnológico de Roque

***Alejandro Guzmán Zazueta***

Instituto Tecnológico de Roque

***María del Socorro Ríos Castro***

Instituto Tecnológico de Roque

***Francisco Gutiérrez Vera***

Instituto Tecnológico de Celaya

*francisco.gutierrez@itcelaya.edu.mx*

## Resumen

Los alumnos de los primeros semestres del Instituto Tecnológico de Roque, deben realizar alguna actividad extraescolar, además de las materias marcadas por su programa académico. Dentro de las actividades extraescolares que se ofrecen a los alumnos, están las cívicas, culturales y deportivas. Las actividades deportivas pueden ayudar a los alumnos a mejorar su estado de salud, siempre y cuando sea una actividad adecuada al alumno, ya que existen algunas enfermedades y padecimientos que combinados con una actividad deportiva, pueden llegar a afectar a quien la practique y padezca de esa enfermedad o padecimiento. La educación integral contempla la impartición de Actividades Extraescolares en los Institutos Tecnológicos

del País. Muchas de estas actividades son deportivas, en donde pueden inscribirse los alumnos. Sin embargo, no se toma en cuenta el historial de salud o alergias que los alumnos puedan tener y que la actividad elegida, puede contribuir a mejorar o empeorar los estados de salud de los alumnos. La presente investigación se llevó a cabo para conocer las enfermedades comunes entre los alumnos del Instituto Tecnológico de Roque, y aquellas actividades extraescolares que les permita mejorar su estado de salud. Posteriormente, se muestra la forma en que se aplicó la minería de datos a la información obtenida, para entregar una propuesta de actividades extraescolares adecuadas para las sintomatologías de los alumnos.

**Palabras Clave:** Alergias, enfermedades, minería de datos.

## **Abstract**

*Students of the first semester of the Institute of Technology Roque, school activity should be performed, in addition to the marked by their academic program. Among the extracurricular activities available to students are the Civic, Cultural and Sports. Sports activities can help students improve their health, long sea when a suitable activity to the student, as there are some diseases and conditions that combined with a sport, may affect the practice and who suffer this disease or condition. Integral Education contemplates the provision of extracurricular activities in the Technological Institutes of the country. Many of these sports son, where they can register the students. However, it does not take into account the medical history or allergies that learners can have and the Chosen of that activity, can help improve or worsen health conditions of students. This research was conducted to know the Commons paragraph diseases among students of the Technological Institute of Roque, and those extracurricular activities that enable them to improve their health. Subsequently, an application of data mining the information obtained, to provide a proposal of extracurricular activities adequate to the symptomatology of Students.*

**Keyboards:** *allergy, data mining, disease.*

## **1. Introducción**

En los últimos años, el problema de obesidad y sobrepeso en México, ha venido incrementándose, sobre todo por los hábitos de vida de los mexicanos, tanto alimenticios como de sedentarismo. Es debido a lo anterior que el gobierno ha establecido programas tanto en escuelas como con programas de televisión, en donde se tiene el interés de atacar estos problemas, ya que son estos, los que pueden derivar en otras enfermedades como hipertensión y diabetes.

No es de sorprender que cada vez es más frecuente escuchar que personas cada vez más jóvenes, presentan problemas de salud como presión arterial alta, o diabetes.

El modelo educativo en México a nivel profesional, está basado en el Modelo de Competencias, en donde el alumno debe aprobar 5 créditos con actividades complementarias a las que realice en sus materias. Dentro de esos 5 créditos, se encuentran contemplados 2 para que el alumno realice actividades extraescolares, las cuales están clasificadas en deportivas, culturales y cívicas. Dentro de las actividades extraescolares, existen fútbol femenino y varonil, básquetbol y natación. Todas estas actividades requieren de esfuerzo físico y algunas de ellas, el contacto con polvo. Si por ejemplo, un alumno padece de asma, este no podrá realizar actividades como fútbol y béisbol, dado que son actividades que tienen contacto con el polvo; y sin embargo, si podrá realizar la actividad de natación, e inclusive le ayudará a contra restar su problema.

Es por esa razón, que se pretende realizar el presente proyecto, para desarrollar un prototipo que proponga a los alumnos, de acuerdo a sus historiales de salud, las actividades extraescolares adecuadas para ellos.

## **2. Métodos**

Para el presente trabajo se hizo una investigación de campo, la cual consistió en acudir con especialistas y con los encargados de impartir las actividades extraescolares, sobre las enfermedades que presentaban los alumnos del Instituto Tecnológico de Roque, así como su relación con actividades deportivas, esto considerando las enfermedades que se registraron en las encuestas aplicadas a los alumnos del Instituto Tecnológico de Roque. Además se hizo una investigación en libros y artículos científicos sobre las mismas. Para desarrollar un poco más este punto, cabe señalar que se acudió con los siguientes especialistas: Cardiólogo, Neumólogo, Médico Internista, Oftalmólogo y Otorrinolaringólogo.

### **2.1 Encuesta a alumnos**

Se aplicó una encuesta a 300 alumnos al azar en el Instituto Tecnológico de Roque, de las carreras de Agronomía, Tics, Industrias Alimentarias, Gestión Empresarial, Hidrología e Innovación Agrícola Sustentable. A los resultados de estas encuestas se les aplicó minería de datos, para su análisis. Las enfermedades obtenidas de estas encuestas fueron hipertensión, diabetes, asma, hepatitis, estrés, anemia y sinusitis. Los impedimentos encontrados son: uso de lentes y lesiones temporales. Las enfermedades e impedimentos de las encuestas serán investigadas en libros, artículos científicos, con especialistas y con los maestros. Además de lo anterior, los docentes manifestaron que los alumnos no hacen caso de sus enfermedades para la realización de su actividad extraescolar.

### **2.2 Limpieza de datos erróneos e inconsistentes de las encuestas**

Primeramente se revisó que las encuestas no tuvieran campos incorrectos, inconsistentes o faltantes, las encuestas que tuvieran al menos un aspecto de los mencionados, se descartaron, los campos que no fueron afectados a estas condiciones fueron de aspecto laboral y actividad extraescolar preferida por el alumno, ya que estos campos no se utilizaron en la aplicación de la minería de datos, por tanto, no son datos

importantes. Las encuestas que no tuvieron alguno de los aspectos mencionados se pasaron a la base de datos. El número de encuestas libres de errores fue de 230.

### **2.3 Clasificación de las actividades extraescolares, con respecto a las afecciones**

Los maestros clasifican las actividades deportivas como fútbol, fútbol americano, baseball, taekwondo, voleibol, atletismo, natación y basquetbol en tres grupos:

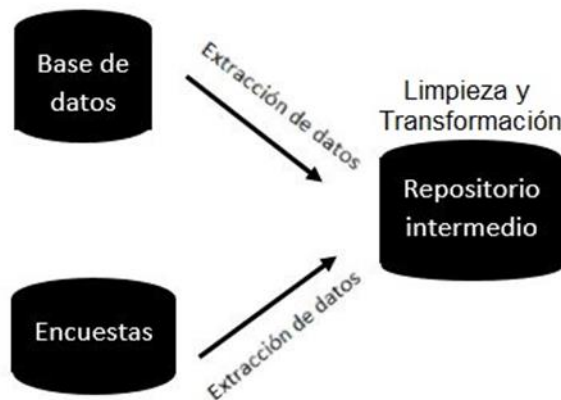
- 1.- Bajo impacto: En este grupo están los alumnos que padecen enfermedades (enfermedades que consideran que el alumno no puede realizar plenamente), consiste en que los alumnos no realicen ejercicios físicos en exceso ni normales, los alumnos realizan ejercicios físicos menores como caminar, saltar, etc. Por un periodo de 10 a 60 minutos según sea el caso, además de sus ejercicios de estiramiento normales y ejercicios de calentamiento menores.
- 2.- Medio impacto: En este grupo están las personas que sufren y no sufren de enfermedad (enfermedades que consideran los maestros que pueden realizarse normalmente) pero padecen de sedentarismo, es decir, alumnos que no están en forma. Consiste en que los alumnos empiecen a realizar ejercicios físicos normales como trotar, correr, sentadillas, lagartijas, etc. Consta también de un periodo de estiramiento alrededor de 5 minutos y uno de calentamiento de alrededor de 10 minutos. El periodo de la actividad física es de 30 a 60 minutos. Este grupo trata de formar alumnos que tengan gran condición física.
- 3.- Alto impacto: En este grupo se encuentran alumnos que no sufren de enfermedad ni de sedentarismo. Los alumnos de este grupo pueden realizar grandes esfuerzos físicos por un periodo mayor a 60 minutos. Para este grupo igualmente se requiere un periodo de estiramiento de alrededor de 5 minutos y uno de calentamiento de alrededor de 10 minutos.

## 2.4 Aplicación de la minería de datos

### 2.4.1 Repositorio intermedio de datos

Para realizar el proceso de minería de datos se necesita primeramente aplicar el sistema ETL (Extraction, Transformation, Load), este sistema pretende extraer, transformar y cargar los datos de la base de datos (fuentes internas) y de las encuestas (fuentes externas) a un repositorio intermedio antes de pasarse al almacén de datos, el propósito de esto es limpiar y transformar los datos que se encuentran en el repositorio intermedio de datos.

En la figura 1 se muestra las fuentes internas y externas utilizadas, las cuales pasaron a formar parte del repositorio intermedio de datos.



**Fig.1. Aplicación del sistema ETL**

Una vez hecho esto, se obtiene una base de datos con datos erróneos y redundancia como se aprecia en la figura 2. Nótese que hay datos como en sexo que contienen “NA”, lo cual significa “No Aplica”, esto es debido a que a esos alumnos no se les aplicó la encuesta. Otro dato importante es que hay datos como en peso, sexo, estatura y abdomen que contienen 99 como valor, igualmente, esto es porque a esos alumnos no se les aplicó la encuesta, este valor tan solo sirve de guía.

controll	nombre	semestre	periodo
10030033	ARROYO MEDINA ANDREA GUADALUPE	8	Enero-Junio
10030291	MOYA MEJIA JUAN ANTONIO	8	Enero-Junio
10940206	ORDOEZ CASILLAS MARITZA SELENE	8	Enero-Junio

idcarrera	clavea	localidad	anno	estado	controll
Industrias A.	Football	ROQUE	2010	IMPRESO	10030033
Industrias A.	Football	ROQUE	2010	IMPRESO	10030291
Agronomía	Football	ROQUE	2010	NO IMPRESO	10940206

enfermedad	enfermedad_h	impedimento	alergia	peso	estatura	edad
Ninguna	Ninguna	Ninguno	Ninguna	77	1.78	23
NA	NA	NA	NA	99	99	99
Diabetes	Ninguna	Ninguno	Ninguna	65	1.65	22

sexo	abdomen
Hombre	70
NA	99
Mujer	65

**Fig. 2. Ejemplo de la información contenida en el repositorio intermedio de datos**

#### 2.4.2 Limpieza de los datos

Para realizar este proceso simplemente se verificó que por medio de consultas, la información del repositorio de datos (encuestas) coincidiera con la información de los alumnos en la base de datos, es decir, verificar que los alumnos de las encuestas existan en la base de datos, que sean de la carrera que dicen ser y que su sexo sea el correspondiente al de la base de datos. Cualquier error en alguna de estas condiciones, el dato se eliminaba. Los datos libres de errores pasan a ser transformados para posteriormente ser depositados en la base de datos y poder aplicarles algún algoritmo de minería de datos.

#### 2.4.3 Transformación de los datos

Los datos libres de errores pasaron a ser transformados por algún método para facilitar la aplicación de minería de datos. En el presente proyecto se aplicó el método de

numerización, el cual consiste en asignar números en vez de letras para facilitar su manipulación. Una vez hecho esto, los datos quedaron como se muestra en la figura 3.

controll	nombre	semestre	periodo	idcarrera
10030033	ARROYO MEDINA ANDREA GUADALUPE	8	Enero-Junio	2
10030291	MOYA MEJIA JUAN ANTONIO	8	Enero-Junio	2
10940206	ORDOEZ CASILLAS MARITZA SELENE	8	Enero-Junio	4

clavea	localidad	anno	estado	controll	enfermedad	enfermedad_h
3	ROQUE	2010	IMPRESO	10030033	0	0
3	ROQUE	2010	IMPRESO	10030291	99	99
3	ROQUE	2010	NO IMPRESO	10940206	0	0

impedimento	alergia	peso	estatura	edad	sexo	abdomen
0	0	0	0	0	0	0
99	99	99	99	99	99	99
0	0	0	0	0	0	0

Fig.3. Ejemplo de los datos transformados

#### 2.4.4 Almacén de datos

Una vez terminado el proceso de la limpieza y transformación de datos que se llevó a cabo en el repositorio intermedio de datos, los datos pasaron a ser migrados al almacén de datos donde se aplicó herramientas de minería de datos como se muestra en la figura 4.

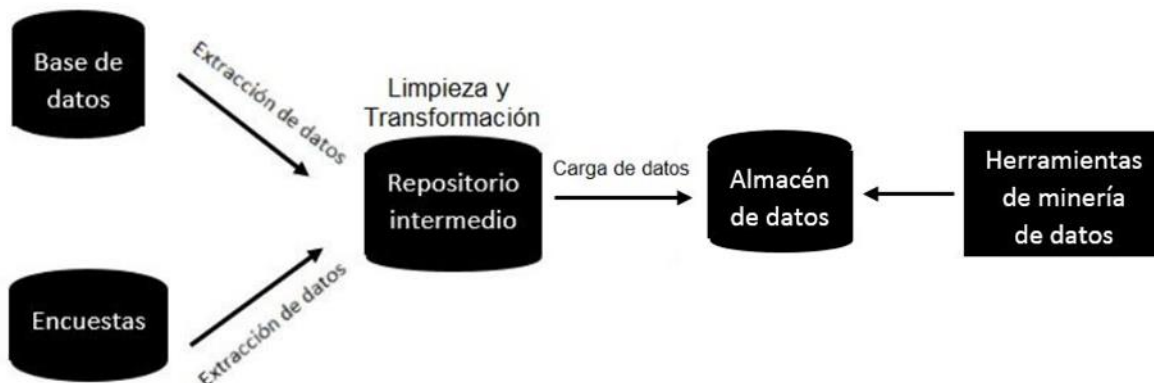
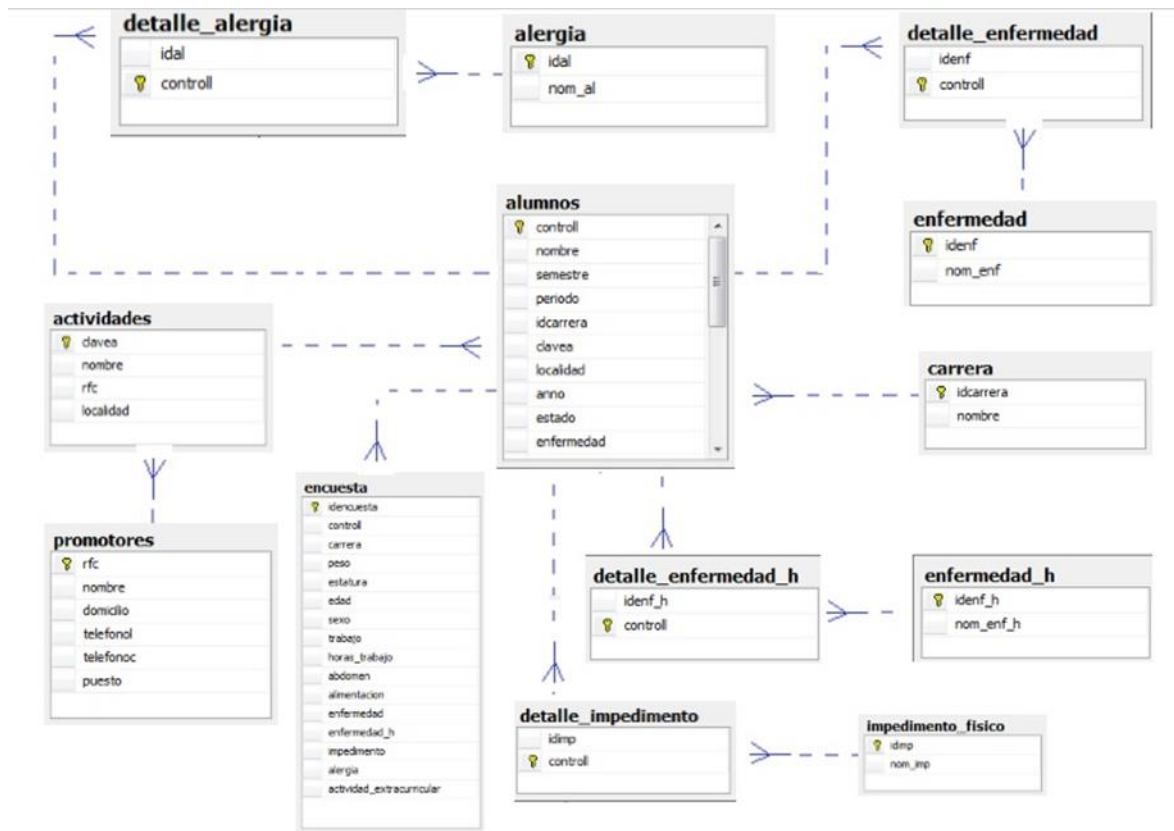


Fig.4. Diagrama del almacén de datos



### 2.4.5 Aplicación del algoritmo de árboles de decisión

Los datos que fueron transformados se depositaron en la base de de Mysql, con lo que el diagrama entidad-relación quedó como se muestra en la figura 5.

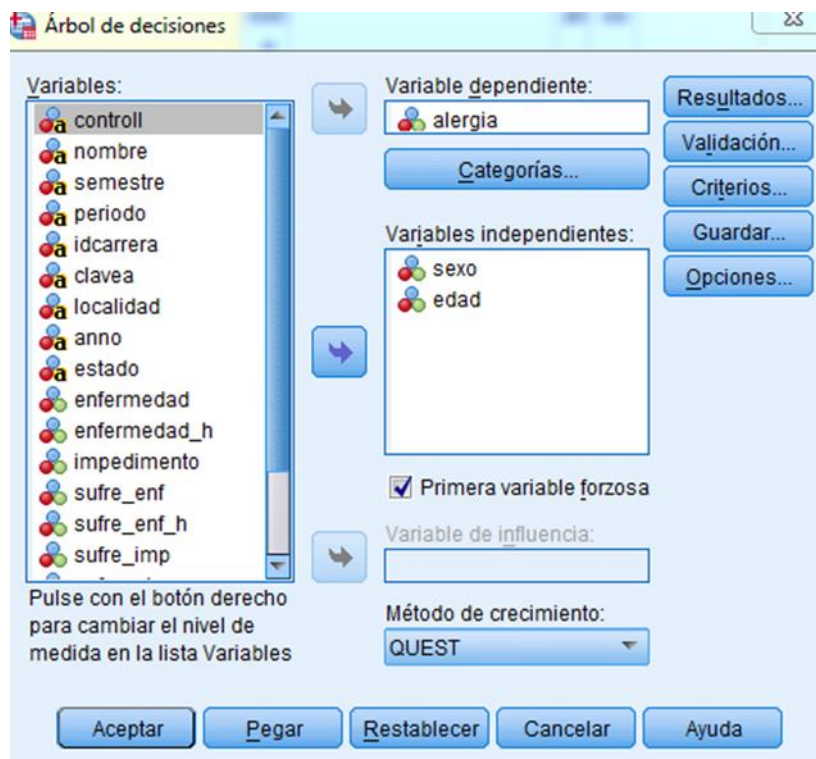


**Fig.5. Diagrama entidad-relación de la base de datos.**

Después la información de la base de datos se pasó a Excel de la siguiente manera:

Primero se debe descargar un software que permita el traspaso de datos de Mysql a Excel, el cual se obtiene en la página de MySQL (MySQL, 2015). Una vez hecho, se debe seguir las instrucciones de instalación las cuales se encuentran en la misma página. Posteriormente, se pasa la información de Excel a IBM para empezar a aplicar la minería de datos.

Al aplicar la minería de datos mediante el método de árboles de decisión se encontró lo siguiente: Para las alergias se muestra los datos insertados en la figura 6.



**Fig.6. Datos insertados para las alergias en el método de árboles de decisión**

A continuación en la figura 7 se muestra el árbol obtenido de los datos insertados, en donde se tiene que las mayores alergias son a los animales, al polvo y a las plantas con un 0.3 %. De lo anterior, la mayoría de los afectados son hombres con un total de 60.1 %. Sin embargo, como puede observarse, las mujeres sufren más de alergias al polvo y animales, los hombres por otra parte sufren más de alergia a las plantas. Se observa que para “ninguna” enfermedad es del 98.6% ya que esa población son a los alumnos que no se le aplicaron la encuesta o no sufren de alergias, se realizó de esta manera debido a que IBM requiere gran cantidad de datos para generar un árbol de decisión.

En cuestión de las enfermedades adquiridas, enfermedades hereditarias, o para algún impedimento físico, se realizó el mismo proceso y se generó el árbol correspondiente a cada situación.

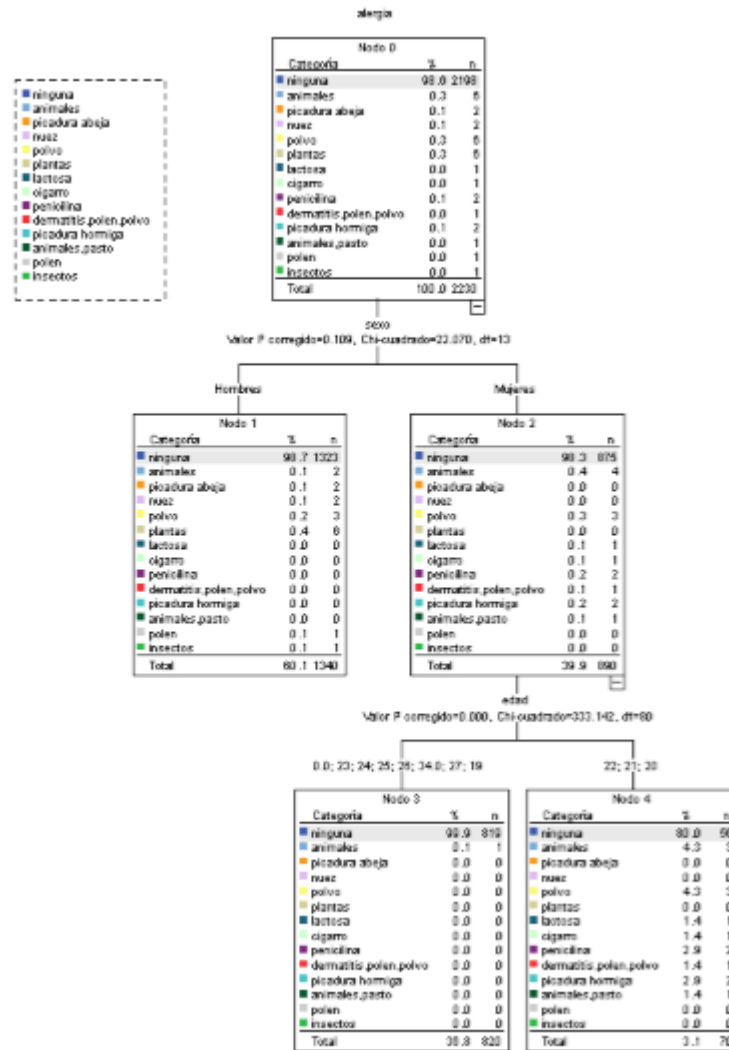
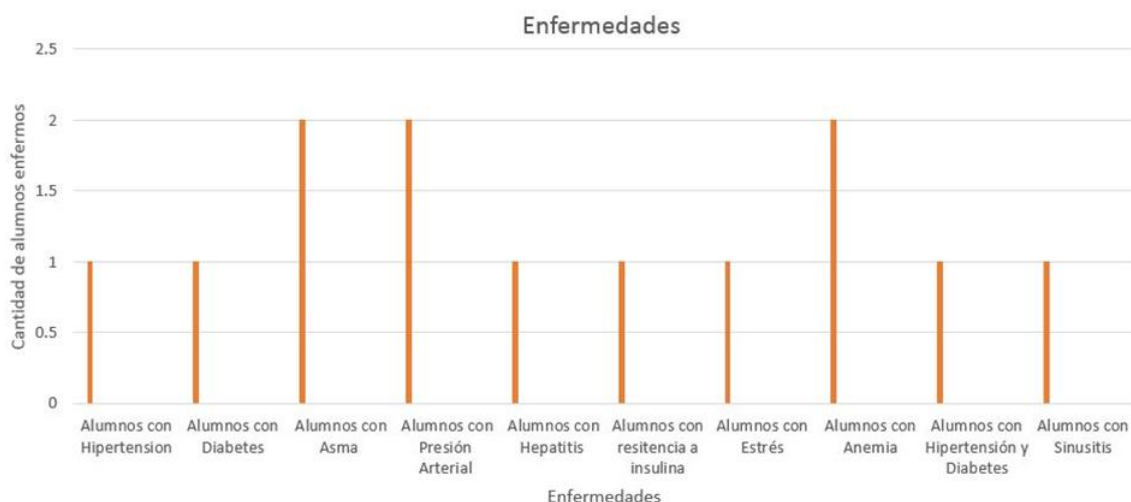


Fig.7. Árbol de decisión de alergias.

### 3. Resultados

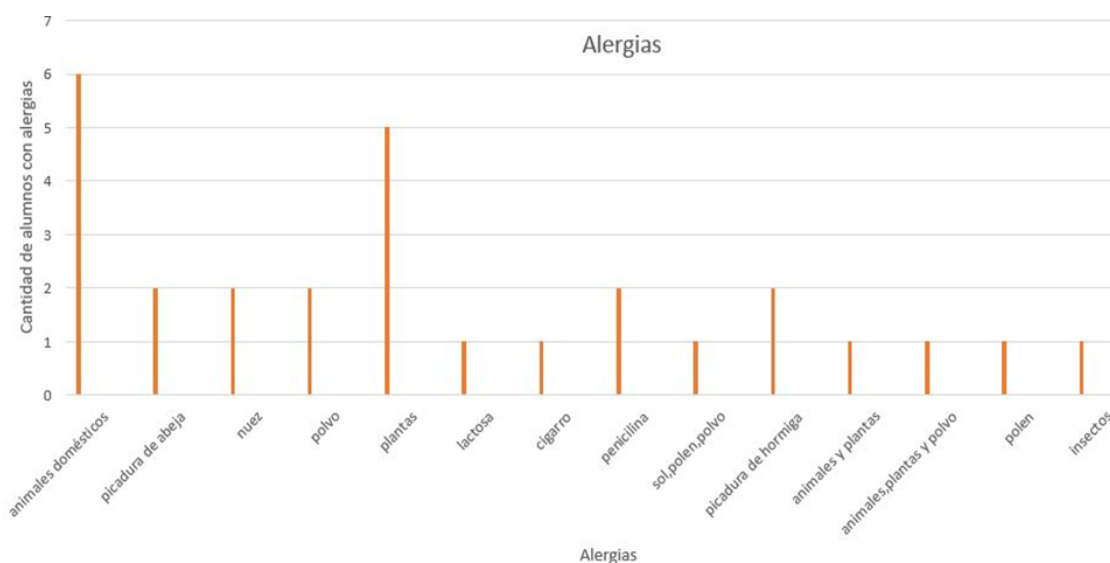
De acuerdo con la muestra que se tomó, se encontró que los alumnos que en su familia padecen hipertensión como enfermedad hereditaria, ningún alumno presentaba alguna enfermedad. Por otro lado, los alumnos que en su familia sufren diabetes como enfermedad hereditaria, uno presentaba la enfermedad de la diabetes, dos la del asma, uno la de presión arterial, uno resistencia a la insulina y uno sinusitis. En cuestión de las enfermedades se encontró que las más destacadas son la diabetes, presión arterial y la anemia como se muestra en la figura 8.



**Fig.8. Enfermedades que presentan los alumnos**

De los cuales todos los alumnos con estas enfermedades están o presentaron football como actividad extraescolar. Por lo antes investigado y por lo que dicen los especialistas, los alumnos que padecen hipertensión, asma y sinusitis están en la actividad extraescolar inadecuada para su salud.

En cuestión de las alergias, se encontró que las enfermedades más relevantes son la alergia a los animales domésticos y a las plantas como se aprecia en la figura 9.

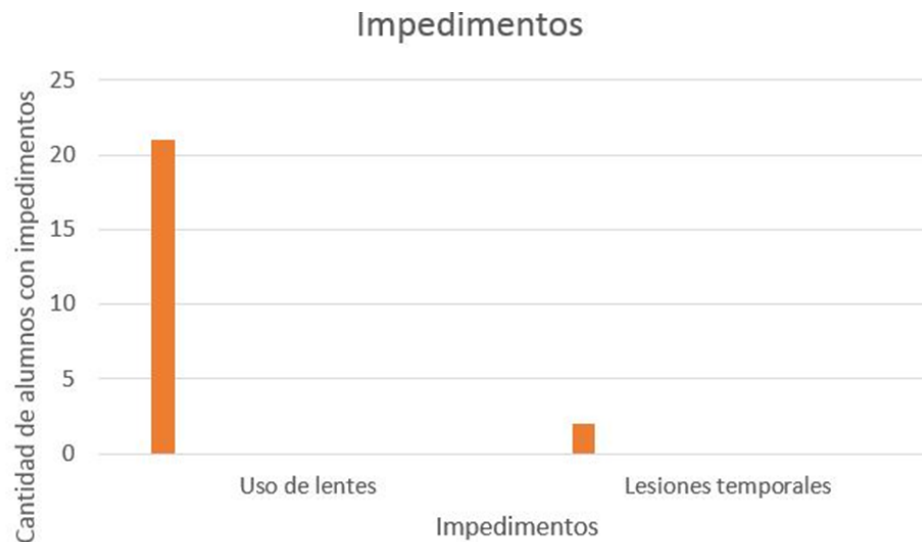


**Fig.9. Alergias que presentan los alumnos**

De los cuales todos los alumnos con estas alergias están o presentaron football como actividad extraescolar.

Con la información que se tiene, consta decir que los alumnos que padecen alergias al polvo, plantas, polen, picadura de abeja, picadura de hormiga e insectos están en una actividad no apta para su salud.

Por último se encontró que el impedimento más influyente es el uso de lentes como se muestra en la figura 10.



**Fig.10. Impedimentos que presentan los alumnos**

## Bibliografía

- [1] José Hernández Orallo, M.José Ramírez Quintana y Cèsar Ferri Ramírez, introducción a la minería de datos, Editorial Pearson, 2004.
- [2] Wilma B. Freire, PH.D, La anemia por deficiencia de hierro: estrategias de la OPS/OMS para combatirla, 1998, [www.scielo.org/scielo.php?pid=S0036-36341998000200012&script=sci\\_arttext](http://www.scielo.org/scielo.php?pid=S0036-36341998000200012&script=sci_arttext)

- [3] V. Plaza Moral, F.J. Álvarez Gutiérrez, P. Casan Clarà, N. Cobos Barroso, A. López Viña, M.A. Llauger Rosselló y J.A. Quintano Jiménez, *Guía Española para el Manejo del Asma*, 2003,  
[www.researchgate.net/profile/Francisco\\_Gutierrez3/publication/224872857\\_Gua\\_Espaola\\_para\\_el\\_manejo\\_del\\_asma\\_\(GEMA\)/links/09e41511e07702937a000000.pdf](http://www.researchgate.net/profile/Francisco_Gutierrez3/publication/224872857_Gua_Espaola_para_el_manejo_del_asma_(GEMA)/links/09e41511e07702937a000000.pdf)
- [4] Ernst P, Spitzer WO, Suissa S, Cockcroft D, Habbick B, Horwitz RI, et al. Risk of fatal and near-fatal asthma in relation to inhaled corticosteroid use, Diciembre 23 de 1992, [jama.jamanetwork.com/article.aspx?articleid=402208](http://jama.jamanetwork.com/article.aspx?articleid=402208)
- [5] Dr. Alberto Maiz Gurruchaga, *CONSECUENCIAS PATOLOGICAS DE LA OBESIDAD: HIPERTENSION ARTERIAL, DIABETES MELLITUS Y DISLIPIDEMIA*, 2002.  
[escuela.med.puc.cl/paginas/publicaciones/Boletin/html/obesidad/obesidad05.html](http://escuela.med.puc.cl/paginas/publicaciones/Boletin/html/obesidad/obesidad05.html)
- [6] *Revista Cubana de Endocrinología* v.13 n.1, Diabetes autoinmune latente del adulto o diabetes tipo 1 de lenta progresión: definición, patogenia, clínica, diagnóstico y tratamiento, Lic. Eduardo Cabrera Rode, 1 Dr. Pedro A. Perich Amador, 2 Dr. Manuel E. Licea, 2002,  
[scielo.sld.cu/scielo.php?script=sci\\_arttext&pid=S1561-29532002000100006](http://scielo.sld.cu/scielo.php?script=sci_arttext&pid=S1561-29532002000100006)
- [7] Dres. Alicia Montano, Raúl Barañano, Beatriz Lageard, et al. Prevalencia de hepatitis A en niños de 2 a 14 años y en población laboral de 18 a 49 años en Montevideo, Uruguay, 2001, , [www.rmu.org.uy/revista/2001v2/art2.pdf](http://www.rmu.org.uy/revista/2001v2/art2.pdf)
- [8] J.A. García Rodríguez, J.E. García Sánchez, M. Gobernado Serrano y J. Mensa Pueyo, *Sociedad Española de Quimioterapia*, Diagnóstico y tratamiento antimicrobiano de la sinusitis. 2003, [www.seq.es/seq/0214-3429/16/2/239.pdf](http://www.seq.es/seq/0214-3429/16/2/239.pdf)
- [9] Dres. Stephen Carroll, Tony Smith, *Enciclopedia de la salud para toda la familia*, editorial Grijalbo, primera edición 1994.
- [10] M. Cortés Berdonces, A. García Martín y M. León Sanz *Nutrición*, Madrid ene.-feb. 2012, Anemia del paciente crítico y quirúrgico, tratamiento con hierro

- intravenoso, revista Hospitalaria vol.27 no.1,  
scielo.isciii.es/scielo.php?pid=S0212-  
16112012000100002&script=sci\_arttext&tIng=en
- [11] Luis Manuel Terán, Michelle Marie Margarete Haselbarth López y David Leonor Quiroz-García, Alergias, pólenes y medio ambiente, 2009,  
[www.anmm.org.mx/GMM/2009/n3/31\\_vol\\_145\\_n3.pdf](http://www.anmm.org.mx/GMM/2009/n3/31_vol_145_n3.pdf)
- [12] MySQL. (Abril, 2015). Página oficial de MySQL.  
<http://dev.mysql.com/downloads/installer/>.
- [13] Instituto Mexicano del Seguro Social, tratamiento de la anemia del prematuro. México, 2012,  
[www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf](http://www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf)  
[www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf](http://www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf)
- [14] Comisión Federal de Mejora Regulatoria, Tabaquismo en México: Análisis y recomendaciones de mejora regulatoria, 2012,  
[www.cofemer.gob.mx/Varios/Adjuntos/14.05.2012/IAPA\\_COFEMER\\_TABAQUISMO\\_EN\\_MEXICO\\_2012.pdf](http://www.cofemer.gob.mx/Varios/Adjuntos/14.05.2012/IAPA_COFEMER_TABAQUISMO_EN_MEXICO_2012.pdf)
- [15] Estudio Nacional de Salud y Envejecimiento en México, boletín de prensa núm. 389/13, Aguascalientes 2013,  
<https://www.google.com.mx/#q=Estudio+Nacional+de+Salud+y+Envejecimiento+en+M%C3%A9xico%2C+bolet%C3%ADn+de+prensa+n%C3%BAm.+389%2F13>
- [16] MedlinePlus, Alérgeno, 12 de mayo del 2015,  
[www.nlm.nih.gov/medlineplus/spanish/ency/article/002229.htm](http://www.nlm.nih.gov/medlineplus/spanish/ency/article/002229.htm)
- [17] Organización Mundial de la Salud, Diabetes, Nota descriptiva N°312, Noviembre de 2014, [www.who.int/mediacentre/factsheets/fs312/es/](http://www.who.int/mediacentre/factsheets/fs312/es/)
- [18] Instituto Mexicano del Seguro Social, Hipertensión arterial, 4 de Marzo del 2015,  
<http://www.imss.gob.mx/salud-en-linea/hipertension-arterial>

- [19] IBM, Manual del usuario del sistema básico de IBM SPSS Statics 20, 2011, [www-947.ibm.com/account/userservices/jsp/login.jsp?persistPage=true&page=/support/entry/myportal/support%3Flnk%3DmsdTS-docu-usen&PD-REFERER=http://www-947.ibm.com/support/entry/myportal/support%3Flnk%3DmsdTS-docu-usen&error=](http://www-947.ibm.com/account/userservices/jsp/login.jsp?persistPage=true&page=/support/entry/myportal/support%3Flnk%3DmsdTS-docu-usen&PD-REFERER=http://www-947.ibm.com/support/entry/myportal/support%3Flnk%3DmsdTS-docu-usen&error=)
- [20] Dr. Alejandro Flores, Clínica de Homeopatía, alérgenos más comunes 2015, [www.homeopatiaflores.com/index.php/areas-de-impacto/alergias/alergenos-comunes](http://www.homeopatiaflores.com/index.php/areas-de-impacto/alergias/alergenos-comunes)
- [21] Microsoft.com, Guía de inicio rápido de Excel 2013. <https://support.office.com/es-mx/article/Gu%C3%ADa-de-inicio-r%C3%A1pido-de-Excel-2013-4337fe93-0690-47cb-89a1-8424d5fdd886?CorrelationId=9b256871-e087-4150-a66b-65a2a5818b10&ui=es-ES&rs=es-MX&ad=MX>