

Aplicación de la minería de datos para determinar propuestas para actividades deportivas en los alumnos

Claudia Rodríguez Lemus

Instituto Tecnológico de Roque
claulemus@itroque.edu.mx

Blanca Cecilia López Ramírez

Instituto Tecnológico de Roque
claulemus@itroque.edu.mx

Alejandro Guzmán Zazueta

Instituto Tecnológico de Roque
claulemus@itroque.edu.mx

Pablo Alfredo Villagómez Naranjo

Instituto Tecnológico de Roque
claulemus@itroque.edu.mx

Francisco Gutiérrez Vera

Instituto Tecnológico de Celaya
francisco.gutierrez@itcelaya.edu.mx

Resumen

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 establecen los expertos para que puedan realizar de acuerdo a su condición e inclusive que les permitan mejorar su estado de salud. Posteriormente, se muestra la aplicación de la minería de datos a la información obtenida, para entregar una propuesta de actividades extraescolares adecuadas para las sintomatologías de los alumnos.

Palabra(s) Clave(s): Actividades deportivas, enfermedades, minería de datos.

Abstract

Integral education includes the incorporation of extracurricular activities in the Technological Institutes of the country. Many of these activities are sports, where students can register. However, it does not check the medical history or allergies that students may have that the chosen activity can help improve or worsen health conditions of students. This research was conducted to know the in common diseases among students of the Technological Institute of Roque, and those extracurricular activities that establish the experts so they can perform according to their condition and even enabling them to improve their health . Subsequently it is shown applying data mining to information obtained, to submit a proposal suitable for symptomatology students extracurricular activities.

Keywords: *3 a 5 palabras claves en inglés, ordenadas alfabéticamente en letra cursiva.*

1. Introducción

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.

Esta situación puede convertirse en ligeros problemas como falta de aire, hasta la muerte por un paro cardíaco.

El índice de sobre peso y obesidad en el país, es muy alarmante y ha originado que el gobierno establezca programas como “chécate, mídete y muévete”, en donde sugiere que primeramente la persona debe realizarse un chequeo. Dado que la mayoría de los alumnos son jóvenes entre 18 y 25 años, no se considera importante el chequeo inicial; es por esa razón que se realizó el presente proyecto, haciendo uso de la herramienta de minería de datos, para entregar propuestas de actividades a los alumnos, de acuerdo a su historial de salud.

2. Métodos

Primeramente se llevó a cabo una investigación de campo, la cual consistió en acudir con especialistas y los maestros 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. 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. Por circunstancias externas a mis posibilidades, no pude hablar con algún Alergólogo debido a que algunos no se prestaban para atender mis preguntas y otros cobraban por proporcionarme información, debido a estos sucesos, omití información de dicho especialista. Posteriormente se realizó investigación de campo, donde se aplicó una encuesta a 300 alumnos al azar en el Instituto Tecnológico de Roque. 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.

A continuación se describe la metodología que se realizó para encontrar los patrones para la propuesta de actividades extraescolares:

Limpiando 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.

Repositorio intermedio de datos

Para realizar el proceso de minería de datos se aplicó el sistema ETL (Extraction, Transformation, Load); con el que se extrajeron, transformaron y cargaron los datos de la base de datos (fuentes internas) y de las encuestas (fuentes externas) a un repositorio intermedio antes de que se pasaran al almacén de datos. Lo anterior se hizo con el propósito de limpiar y transformar los datos del repositorio intermedio.

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

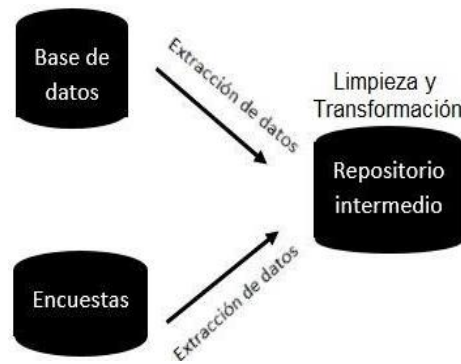


Figura 1 Aplicación del sistema ETL.

Una vez hecho esto, se obtuvieron tablas con datos erróneos y redundancia como se aprecia en la figura 2. Nótese que hay campos 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. Ahora estos datos están listos para el siguiente paso, la limpieza de datos.

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

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

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 cuestas 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 elimina. 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.

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 este proyecto se aplicó el método de numerización, el cual consiste en asignar números en vez de letras para facilitar su manipulación. A continuación se explica la asignación de los valores numéricos:

Se asignaron numeraciones a las carreras, a las actividades deportivas, a las enfermedades, a las alergias y al género de los alumnos. Al final, los datos se transformaron como se muestra en la figura 3.

control	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	control	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

Figura 3 Ejemplo de los datos transformados.

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

las instrucciones de instalación las cuales se encuentran en la misma página. Posteriormente, se pasó la información de Excel a IBM para empezar a aplicar la minería de datos.

En la figura 6 se muestra la manera como se aplicó el método de minería a través de árboles de decisión, tomando como variable dependiente a la enfermedad y como variables independientes al sexo y abdomen.

Después de aplicar los árboles de decisión, se encontró que de las enfermedades mostradas, la más destacada es la hipertensión, de la cual las mujeres que tienen un abdomen entre 60 a 125 sufren más de esta enfermedad (ver figura 6).

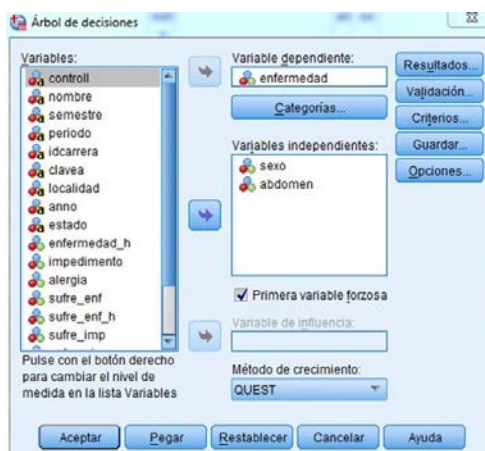


Figura 6 Datos insertados para las alergias en el método de árboles de decisión.

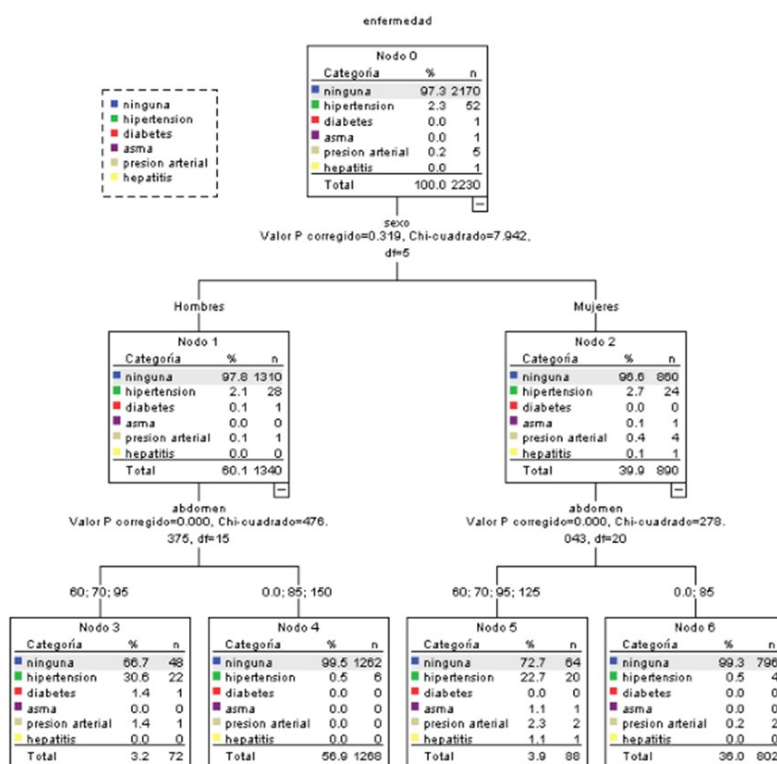


Figura 6 Árbol de enfermedades.

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, 1 presentaba la enfermedad de la diabetes, 2 la del asma, 1 la de presión arterial, 1 resistencia a la insulina y 1 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 7.

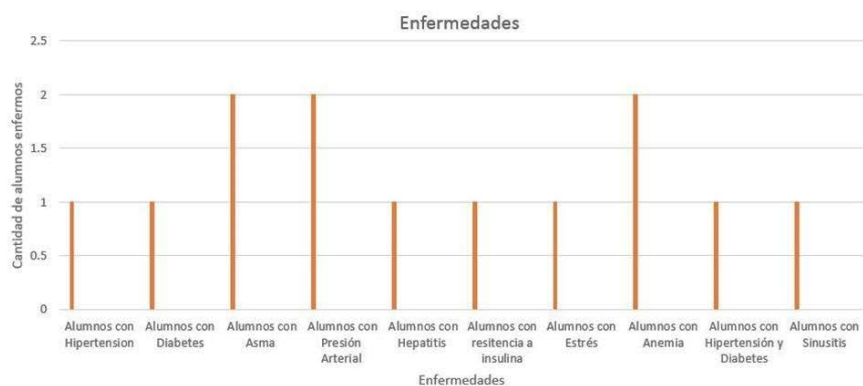


Figura 7 Enfermedades que presentan los alumnos.

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

Con la información que se obtuvo, se puede 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. Estos alumnos presentaron football como actividad extraescolar.

De acuerdo a la información de los especialistas y lo antes investigado, los alumnos que padecen impedimentos al uso de lentes y que tienen una graduación de 8 o más dioptrías están en una actividad extraescolar no apta para ellos.

Cabe señalar que con los resultados obtenidos y de la manera en que los alumnos exponen su salud al realizar una actividad extraescolar no apta para ellos, se hace mención de las actividades extraescolares recomendadas para aquellos alumnos que padecen enfermedades como:

- Hipertensión y presión arterial: Rondalla, Teatro, Danza y Natación.

- Asma: Natación, Tae Kwon Do, Rondalla, Teatro y Danza.
- Sinusitis: Natación, Rondalla, Teatro y Danza.
- Diabetes: Football, Natación, Atletismo, Baseball, Basketball, Volleyball, y Tae kwon Do.
- Hepatitis: Cualquier actividad extraescolar.
- Anemia: Cualquier actividad extraescolar.
- Resistencia a la insulina: Cualquier actividad extraescolar.
- Estrés: Cualquier actividad extraescolar excepto banda de guerra.

Se recomienda actividades extraescolares para aquellos alumnos que padecen alergias como:

- Animales domésticos: Cualquier actividad extraescolar.
- Picadura de abeja y picadura de hormiga: Natación, Danza, Teatro, Rondalla, Banda de guerra, Basketball, Volleyball, Atletismo y Tae Kwon Do.
- Nuez: Cualquier actividad extraescolar
- Polvo, plantas, insectos y polen: Natación, Danza, Teatro, Rondalla, Basketball, Volleyball, Atletismo y Tae Kwon Do.
- Lactosa: Cualquier actividad extraescolar.
- Cigarro: Cualquier actividad extraescolar excepto banda de guerra y rondalla.
- Penicilina: Cualquier actividad extraescolar.

Se recomienda actividades extraescolares para aquellos alumnos que padecen impedimentos físicos como:

- Uso de lentes con 8 dioptrías o más: Danza, Teatro, Rondalla, Banda de guerra, Atletismo y Natación.
- Uso de lentes con menos de 8 dioptrías: Cualquier actividad extraescolar.

4. Discusión

Se hace mención que la investigación de campo en combinación con la investigación documental en artículos científicos y libros, utilizada para este trabajo de investigación fue la adecuada, ya que se pudo combinar información de libros y de especialistas que

influyó para determinar las mejores actividades extraescolares que pudieran practicar los alumnos que presentaran alguna enfermedad, alergia o impedimento.

Con respecto a la herramienta de minería de datos de árboles de decisión, utilizada para determinar la relación que había en cuanto al estado de salud de los alumnos y las enfermedades más influyentes, fue de gran ayuda ya que muestra una amplia visualización de los conjuntos encontrados y el porcentaje que destacan.

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
- [3] OPS/OMS para combatirla, 1998, www.scielosp.org/scielo.php?pid=S003636341998000200012&script=sci_arttext.
- [4] 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).
- [5] 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
- [6] Dr. Alberto Maiz Gurruchaga, Consecuencias Patologicas De La Obesidad: Hipertension Arterial, Diabetes Mellitus Y Dislipidemia, 2002. scuela.med.puc.cl/paginas/publicaciones/Boletin/html/obesidad/obesidad05.html
- [7] 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.

- [8] 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.
- [9] 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.
- [10] Dres. Stephen Carroll, Tony Smith, Enciclopedia de la salud para toda la familia, editorial Grijalbo, primera edición 1994.
- [11] 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&tlng=en.
- [12] Luis Manuel Terán, Michelle Marie Margarete Haselbarth López y David Leonor.
- [13] Quiroz-García, Alergias, pólenes y medio ambiente, 2009, www.anmm.org.mx/GMM/2009/n3/31_vol_145_n3.pdf.
- [14] MySQL. (Abril, 2015). Página oficial de MySQL. <http://dev.mysql.com/downloads/installer/>.
- [15] Instituto Mexicano del Seguro Social, tratamiento de la anemia del prematuro. México, 2012, www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf. www.imss.gob.mx/sites/all/statics/guiasclinicas/540GER.pdf.
- [16] 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
- [17] 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> [19] MedlinePlus, Alérgeno, 12 mayo 2015, www.nlm.nih.gov/medlineplus/spanish/ency/article/002229.htm
- [18] Organización Mundial de la Salud, Diabetes, Nota descriptiva N°312, Noviembre de 2014, www.who.int/mediacentre/factsheets/fs312/es/.

- [19] Instituto Mexicano del Seguro Social, Hipertensión arterial, 4 de Marzo del 2015, <http://www.imss.gob.mx/salud-en-linea/hipertension-arterial>.
- [20] 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://www947.ibm.com/account/userservices/jsp/login.jsp?persistPage=true&page=/support/entry/myportal/support%3Flnk%3DmsdTS-docu-usen&PD-REFERER=) y <http://www947.ibm.com/support/entry/myportal/support%3Flnk%3DmsdTS-docuusen&error=>.
- [21] 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).
- [22] Microsoft.com, Guía de inicio rápido de Excel 2013. <https://support.office.com/es-mx/article/Gu%C3%ADa-de-inicio-r%C3%A1pidode-Excel-2013-4337fe93-0690-47cb-89a1-8424d5fdd886?CorrelationId=9b256871-e087-4150-a66b-65a2a5818b10&ui=esES&rs=es-MX&ad=MX>.