

Aplicación de la minería de datos para la selección de biomateriales en implantes dentales

Application of data mining for the selection of biomaterials in dental implants

Nancy B. Ganz¹, Alicia E. Ares¹, Horacio D. Kuna²

¹Laboratorio de Ciencia de los Materiales, Instituto de Materiales de Misiones (IMAM-CONICET), Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.

e-mail: nancy.bea.ganz@gamil.com, a.e.ares@gmail.com

²Facultad de Ciencias Exactas Químicas y Naturales (FCEQyN), Universidad Nacional de Misiones (UNaM), Félix de Azara 1552, N3300LQH, Posadas, Misiones, Argentina.

RESUMEN

En Argentina, como en otros países del mundo se fabrican biomateriales para diferentes aplicaciones con el objetivo de restaurar las funciones del cuerpo humano, como es el caso de los implantes dentales. Sin embargo, es necesario contar con información suficientemente calificada y accesible sobre: características de los implantes dentales, datos de la fase quirúrgica y condiciones de salud de los pacientes. La carencia de un registro digital de acceso público con datos reales de historias clínicas de pacientes que se han sometido a procesos quirúrgicos para la colocación de implantes dentales, dificulta la tarea de análisis y extracción de conocimiento para los especialistas implantólogos, sobre la relación existente entre las características del implante y las condiciones del paciente. De aquí, surgió la necesidad de crear un registro automatizado que reúna las variables que representan el proceso de colocación de un implante dental, con el objetivo de identificar los factores que contribuyen al éxito o al fracaso de los implantes dentales colocados en la Provincia de Misiones - Argentina, a través de la aplicación de técnicas de minería de datos. En este trabajo se detallan las variables del conjunto de datos *BD_IMPLANTES.csv*, el cual está conformado por cuatro dimensiones, estas son: paciente, implante, fase quirúrgica y seguimiento postoperatorio. El objetivo fue conocer y analizar las variables de mayor importancia o aquellas que influyen sobre el atributo clase. Se observó en este primer relevamiento que una de las variables que aporta ganancia de información significativa es la variable *trat_sup* (tratamiento de superficie del implante), por lo cual puede ser una determinante para el proceso de oseointegración tejido/implante.

Palabras clave: Minería de Datos, Ganancia de Información, Biomateriales, Implantes Dentales.

ABSTRACT

In Argentina, as in other countries of the world, biomaterials are manufactured for different applications with the objective of restoring the functions of the human body, as is the case with dental implants. However, it is necessary to have sufficiently qualified and accessible information on: characteristics of dental implants, data from the surgical phase and health conditions of patients. The lack of a publicly accessible digital record with real data from medical records of patients who have undergone surgical processes for the placement of dental implants, makes it difficult for implant specialists to analyze and extract knowledge about the relationship between the characteristics of the implant and the conditions of the patient. From here, the need arose to create an automated register that gathers the variables that represent the process of placement of a dental implant, in order to identify the factors that contribute to the success or failure of dental implants placed in the Province of Misiones - Argentina, through the application of data mining techniques. This work details the variables of the data set *BD_IMPLANTES.csv*, which is made up of four dimensions, these are: patient, implant, surgical phase and postoperative follow-up. The objective was to know and analyze the most important variables or those that influence the class attribute. It was observed in this first survey that one of the variables that provides significant information gain is the variable *trat_sup* (implant surface treatment), so it can be a determinant for the process of tissue/implant osseointegration.

Keywords: Data Mining, Information Gain, Biomaterials, Dental Implants.

1. INTRODUCCIÓN

Actualmente, el campo de la minería de datos [1] ha tenido considerables avances respecto a la aplicación y desarrollo de técnicas en el sector de la salud, para la predicción de enfermedades y para la toma de decisiones en base al análisis de grandes cantidades de datos. Existen ciertos trabajos que aplican métodos para la predicción del éxito de los implantes dentales, como TAMEZ *et al.* [2] mostraron un análisis estadístico de Regresión Logística Múltiple para determinar los factores que influyen en el éxito de los implantes dentales. DOMÍNGUEZ *et al.* [3] realizaron un estudio donde determinan si existe relación entre los fracasos de los implantes dentales y las enfermedades sistémicas (concretamente sobre la osteoporosis, hipertensión, diabetes e hipotiroidismo), en una población de pacientes sometidos a cirugía de implantes dentales en el hospital San José de Santiago de Chile. En el trabajo de OLIVEIRA *et al.* [4] presentan un análisis comparativo de tres técnicas de aprendizaje automático: Máquina de Vector Soporte (SVM), Máquina de Vector Soporte Ponderado y Redes Neuronales RBF constructivas (RBF-DDA) con parámetro de selección, para la predicción del éxito de los implantes dentales. Si bien estos trabajos utilizan conjuntos de datos de implantes dentales, no centran su atención al biomaterial (tratamiento de superficie), es por esto que este trabajo aborda conjuntamente el estudio de las características del implante propiamente dicho, ligado a los rasgos y condiciones de salud de los pacientes.

Asimismo, para documentar proyectos de Minería de Datos, existen procesos diseñados para guiar al usuario a través de una sucesión formal de pasos. Entre las metodologías más importantes encontramos: SEMMA, CRISP-DM [5] y P3TQ. La metodología CRISP-DM cuenta con 6 fases que forman un proceso iterativo-incremental, las cuales son: comprensión del negocio, comprensión de los datos, preparación de los datos, modelado, evaluación y despliegue.

Este trabajo de investigación permitió la creación de un registro digital con datos reales de historias clínicas de pacientes que se han sometido al proceso quirúrgico de colocación de implantes dentales en las localidades más características de la provincia de Misiones. Se trabajó en conjunto con expertos en el área de Minería de Datos, Biomateriales, Especialista en Implantología y Rehabilitación Compleja, Especialista en Patología Bucal e Implantología Oral, así como en Estadística.

Para comenzar con el proceso de extracción de conocimiento, los datos utilizados en este trabajo fueron obtenidos manualmente a través del diseño y corroboración de un formulario por parte de los expertos en el área. Así como, sucesivas entrevistas y charlas con más de 30 especialistas en implantología oral, con el fin de determinar cuáles eran los datos de importancia a la hora de considerar la colocación de un implante dental.

2. MATERIALES Y MÉTODOS

En base a la metodología CRISP-DM, la fase 1 de comprensión del problema abarcó: la definición del objetivo del presente trabajo, siendo la determinación de los factores que contribuyen al éxito de los implantes dentales en la provincia de Misiones, Argentina; se evaluó la situación de la carencia de un registro digital y estandarizado sobre casos de implantes dentales colocados; se precisó el plan de trabajo, que consistió en el relevamiento bibliográfico, investigación de la situación actual de la implantología en la provincia, selección de los expertos en el área, definición de la metodología a implementar, diseño del formulario en base a los objetivos, inicio del proceso de recolección de datos a través del formulario confeccionado y preparación de los datos (selección, limpieza, conversión y formato).

En la fase 2 de comprensión de los datos se plasmó la recolección de los datos iniciales, donde se observó que los datos se agrupan en 4 dimensiones:

Datos del Paciente: referidos a los antecedentes y condiciones médicas de los pacientes a la hora de la intervención.

Datos del Implante: características del implante utilizado por el especialista implantólogo.

Datos de la Fase Quirúrgica: referidos al procesamiento de intervención quirúrgica y mejoramiento del lecho óseo del paciente.

Datos del Seguimiento Postoperatorio: datos del resultado del proceso de colocación del implante, es decir si el proceso de oseointegración tejido / implante tuvo éxito o fracasó.

Estas 4 dimensiones se encuentran alojadas en un solo archivo de datos, denominado “*BD_IMPLANTES.csv*”, está compuesto por un total de 32 variables, una variable target o label (SEGUI_POSTOP) y 1050 registros o filas, las cuales representan casos de implantes colocados. La descripción detallada se proporciona en la Tabla 1.

Tabla 1: Detalles del conjunto de datos de implantes dentales examinado.

Base de Datos	<i>BD_IMPLANTES.csv</i>
Nro. de clases	1 atributo binario
Nro. de características	31
Nro. de ejemplos	1050
Muestras Negativas	73 fracasos
Muestras Positivas	977 éxitos

Finalizada la fase de exploración inicial de los datos, se pudo afirmar que los datos son completos, cubren los casos requeridos para cumplir con los objetivos, no contienen errores, no se encontraron valores fuera de rango y nulos.

En la fase 3 de preparación de los datos se realizó la transformación y reducción de las variables que no aportaban ganancia de información al estudio de caso, con el fin de construir un conjunto de datos final para iniciar con el diseño de un procedimiento que utilice técnicas de minería de datos, e ir ajustando en función de las necesidades que se presenten. Por ejemplo, se transformó la variable edad en rangos de edades, las enfermedades por tipos, los medicamentos por la tipificación de las drogas, la fecha de intervención por la estación del año, para la variable número dental se generó otra que se corresponde con las zonas de cada cuadrante de la boca, se combinó las variables relacionadas al implante para obtener una nueva variable con datos del tratamiento de su superficie, y así distintas composiciones (siempre teniendo en cuenta la opinión del experto) para ir experimentando y definiendo relaciones entre las variables. Seguidamente, en la tabla 2 se detallan los atributos que componen cada una de las dimensiones del conjunto de datos final, con una breve descripción.

Tabla 2: Dimensiones del conjunto de datos.

Dimensión I - Datos del Paciente	
Característica	Descripción
<i>rango_edad</i>	Rango de edad en años cumplidos por el paciente.
<i>genero</i>	Sexo del paciente.
<i>profesion</i>	Ocupación en el que se desempeña el paciente.
<i>o_s</i>	Determina si la intervención implantología es por seguro médico o de manera particular.
<i>antecedente</i>	Enfermedad que padece el paciente.
<i>fuma</i>	Este campo determina si el paciente es fumador.
<i>alcoholismo</i>	Define si el paciente es alcohólico.
<i>periodontitis</i>	Determina el grado de afección que tiene el tejido o hueso alrededor de una o varias piezas dentarias.
<i>ingest_med</i>	Determina si el paciente se encuentra tomando algún tipo de medicamento.
<i>alergia</i>	Establece si el paciente tiene alguna alergia.
Dimensión II - Datos del Implante	
<i>trat_sup</i>	Este campo brinda información sobre el tratamiento de superficie del implante. Considerando: T1: Blasting + doble grabado ácido; T2: Blasting + grabado ácido + tratamiento térmico; T3: Blasting + doble grabado ácido + depósitos de nano cristales de fosfato cálcico; T4: Grabado bi-ácido; T5: Blasting + grabado ácido; T6: Blasting con fosfato tricálcico + grabado ácido o T7: Tratamiento de Electroquímico.
<i>diseño</i>	Define la forma o macrogeometría del implante.
<i>longitud</i>	Concepto métrico en milímetros que define el largo o extensión del implante.
<i>diametro</i>	Concepto métrico en milímetros que define el grosor o espesor del implante.
<i>conexion</i>	Representa la forma de conexión entre el pilar y el implante, definiendo la existencia o ausencia de una figura geométrica que se extiende por sobre la superficie de la corana del implante.
<i>procedencia</i>	Determina el origen del implante. El cual puede ser nacional o importado.
Dimensión III - Datos de la Fase Quirúrgica	
<i>estacion_anio</i>	Indica la estación del año en el que el paciente se realizó la intervención quirúrgica.
<i>zona_paciente</i>	Indica la zona en la cual reside el paciente.
<i>registro</i>	Permite comprobar si el odontólogo pertenece al “Registro Provincial de Profesionales que practican Cirugía Buco maxilofacial, Implantología, Periodoncia y Manipulación de Tejidos”. Este registro permite precisar si el especialista posee dos años de experiencia y 200 horas mínimas en cursos de especialización.

<i>pieza_dentaria</i>	Este campo identifica la localización o el cuadrante del implante en la boca del paciente.
<i>protocolo_carga</i>	Hace referencia al tiempo transcurrido entre la extracción de la pieza dentaria y la colocación del implante dental. Puede ser inmediato (24-48 hs.), temprano (posterior a las 48 hs. y dentro del mes) o tardío (posterior al mes).
<i>exodoncia</i>	Establece si se produjo un acto quirúrgico en el que se extrajo el diente o una parte remanente del mismo que estaba alojado en el alveolo.
<i>expan_osea</i>	Determina si se realizó alguna técnica para incrementar la anchura del hueso mandibular antes de la colocación del implante dental.
<i>elev_seno_maxilar</i>	Determina si se realizó alguna técnica para incrementar la altura ósea en el maxilar superior (al nivel de los molares y premolares) para la colocación del implante dental.
<i>reg_tej_duros</i>	Determina si se regeneró los tejidos duros (hueso) donde se colocó el implante.
<i>reg_tej_blandos</i>	Determina si se regenero los tejidos blandos (encía) donde se colocó el implante.
<i>proc_adicional</i>	Establece si el paciente pasó por una instancia de tratamiento previo de preparación de lecho óseo.
<i>tiem_coloc</i>	Hace referencia al tiempo transcurrido entre la colocación del implante y la carga de la corona. Puede ser inmediato (contiguo a la colocación del implante), temprano (dentro de los 3 meses) o tardío (posterior a los 3 meses).
<i>tipo_hueso</i>	Define la calidad o tipo de hueso del paciente en la zona a ser intervenida. Puede ser tipo I (hueso muy denso), tipo II (hueso trabecular denso), tipo III (hueso trabeculado más esponjoso) o tipo IV (hueso con trabéculas grandes y débiles).
<i>indic_protésica</i>	Determina la complejidad de la cirugía y la carga que tendrá el implante. Puede ser pieza unitaria, puente, prótesis completa fija o removible.
<i>comp_quirurgicas</i>	Este concepto determina si se produjo alguna complicación en la intervención quirúrgica.
Dimensión IV - Datos del Seguimiento Postoperatorio	
<i>seguí_postop</i>	Determina si el implante tuvo éxito en el proceso de oseointegración o el proceso fracaso.
<i>observaciones</i>	Establece porque razón el implante fracaso o simplemente que rasgos se notaron luego del proceso de curación.

3. RESULTADOS

A continuación, en la figura 1 se presenta el peso o nivel de importancia (eje y) obtenido a través del algoritmo de ganancia de información [6, 7] de las variables ajustadas (eje x) en función de la variable label (*seguí_postop*), es decir en base al éxito o fracaso del proceso de integración tejido / implante:

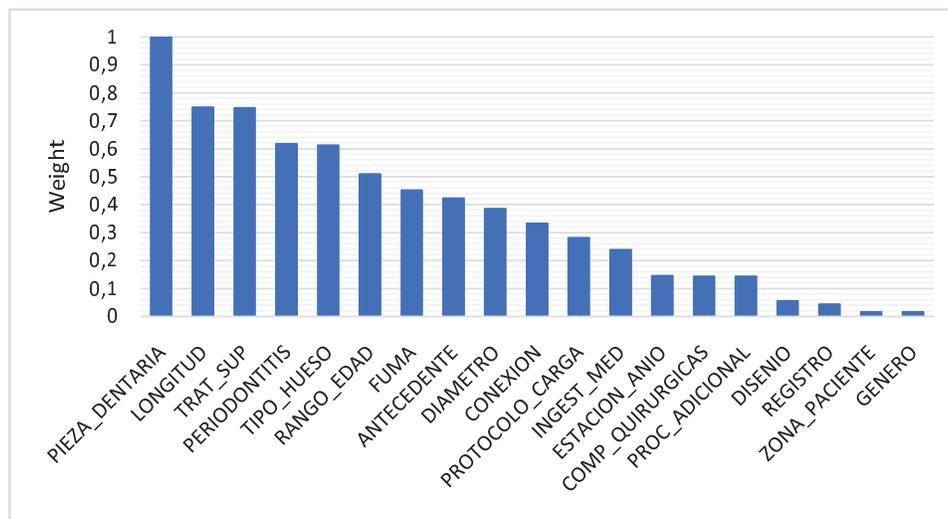


Figura 1: Ganancia de Información.

4. DISCUSIÓN

A la hora de analizar las características o variables de un conjunto de datos, el proceso debe comprender el

análisis y la prueba de varios métodos de selección de características, para no sesgar la decisión en función de los resultados de un sólo método. Es por esto que es conveniente realizar una combinación de métodos para tener una visión más abarcativa y no tan estricta a una sola técnica. Pero el objetivo de este trabajo fue lograr un primer acercamiento a los datos para estudiarlos y comprenderlos, mediante la realización de un diccionario de datos (tabla 2) que plasme las especificaciones de cada variable. Además, se logró confeccionar un conjunto de datos limpios y preparados para la posterior aplicación de técnicas de minería de datos, para la identificación de los factores que contribuyen al éxito o al fracaso de los implantes dentales.

5. CONCLUSIONES

En base a los resultados obtenidos de aplicar el algoritmo de ganancia de información, pudimos observar que las variables que tiene mayor influencia sobre la variable target, en base al conjunto de datos extraído de historias clínicas de pacientes de la Provincia de Misiones, Argentina, fueron: pieza dentaria, longitud, tratamiento de superficie, periodontitis, tipo de hueso, rango de edad, fuma y antecedentes. Estas variables obtuvieron un valor mayor a 0,4 de nivel de importancia respecto a todas las variables del conjunto de datos.

Estos resultados fueron validados mediante consultas y opiniones con los expertos, los cuales coincidieron y remarcaron que los factores: pieza dentaria, antecedentes, periodontitis, tipo de hueso, edad, característica del implante (como longitud y diámetro), protocolo de carga y mejoramiento del lecho óseo (procedimiento adicional), son determinantes a la hora de planificar una cirugía oral para la colocación de un implante dental.

Mientras que para las variables: estación del año, procedencia del paciente y tratamiento de superficie, los especialistas han reflexionado que son cuestiones interesantes que requieren de un estudio en detalle.

Así mismo, se apreció que las características del implante (longitud, diámetro, diseño, conexión y tratamiento de superficie) son determinantes en función de los datos examinados. Lo que hace necesario un estudio en mayor detalle de la tipificación de la variable *trat_sup*.

En referencia a estas opiniones, se validaron los factores detectados en el conjunto de datos de implantes dentales utilizado.

AGRADECIMIENTOS

Agradecemos al Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) por el financiamiento a través de una beca doctoral y a los especialistas implantólogos que colaboraron en la construcción de la base de datos de implantes dentales.

BIBLIOGRAFÍA

- [1] FAYYAD U., PIATETSKY-SHAPIRO G., SMYTH P., "The KDD Process for Extracting Useful Knowledge from Volumes of Data", *Communications of the ACM*, v. 39, n. 11, pp. 27–34, 1996.
- [2] TAMEZJ. E. B., ZILLI F. N., FANDIÑO L. A., GUIZARJ. M., "Factores relacionados con el éxito o el fracaso de los implantes dentales colocados en la especialidad de Prostodoncia e Implantología en la Universidad de La Salle Bajío", *Revista Española de Cirugía Oral y Maxilofacial*, v. 286, pp. 1–9, 2016.
- [3] DOMÍNGUEZ J., ACUÑA J., ROJAS M., BAHAMONDES J., MATUS S., "Study of association between systemic diseases and dental implant failure", *Revista Clínica de Periodoncia, Implantología y Rehabilitación Oral*, v. 6, n. 1, pp. 9–13, 2013.
- [4] OLIVEIRAA. L. I., BALDISSEROTTO C., BALDISSEROTTO J., "A comparative study on machine learning techniques for prediction of success of dental implants", *Mexican International Conference on Artificial Intelligence*, v. 3789, pp. 939–948, 2005.
- [5] CHAPMAN, P., CLINTON, J., KERBER, R., KHAHAZA, T., REINARTZ, T., SHEARER, C., WIRTH, R. P., "CRISP-DM 1.0: Step-by-step data mining guide", 2000.
- [6] SHANNON C. E., "A Mathematical Theory of Communication", *The Bell System Technical Journal*, v. 27, n. 3, pp. 379–423, 1948.
- [7] ABE N., KUDO M., "Entropy Criterion for Classifier-Independent Feature Selection", *International Conference on Knowledge-Based and Intelligent Information and Engineering Systems*, n. 1, pp. 689–695, 2005.