



APLICACIÓN DE LA MINERÍA DE DATOS PARA LA SELECCIÓN DE BIOMATERIALES EN IMPLANTES DENTALES

Nancy B. Ganz^{(1,2)*}, Alicia E. Ares^(1,2) y Horacio D. Kuna⁽²⁾

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

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

* Correo Electrónico: nancy.bea.ganz@gmail.com

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. No obstante, 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. En este trabajo, se detallan las variables que componen el registro digital de datos reales de historias clínicas de pacientes que se han sometido al proceso quirúrgico de colocación de implantes dentales en la Provincia de Misiones, Argentina.

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 en [2] se muestra un análisis estadístico de Regresión Logística Múltiple para determinar los factores que influyen en el éxito de los implantes dentales. En [3] se presenta un estudio donde determina 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 [4] se presenta 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 datasets 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 (datasets) 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.

En base a la metodología CRISP-DM, la *fase 1 de comprensión del problema* abarcó la definición del objetivo del presente trabajo, el cual es 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 fijó el objetivo de minería de datos, siendo el diseño de un procedimiento a través de una metodología híbrida; 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 de minería de datos a aplicar, 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 formateo). 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:** concernientes a las 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:** alusivos al 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 34 variables, una variable target o label (SEGUI_POSTOP) y 1050 registros o filas, las cuales representan casos de implantes colocados en la provincia de Misiones. Existen variables cuantitativas (*edad, longitud y diámetro*) y variables cualitativas categóricas (*antecedentes, fuma, periodontitis, diseño, pieza dentaria, entre otras*).

Seguidamente, en la tabla 1 se detalla el contenido (columnas) del datasets correspondiente a cada dimensión:

Dimensión	Columnas
Paciente	<i>Edad, género, ocupación, obra social, enfermedad, fuma, alcoholismo, periodontitis, riesgo periodontal, ingesta de medicamentos, alergia, lugar de procedencia.</i>
Implante	<i>Material, marca, diseño, longitud, diámetro, conexión, procedencia.</i>
Fase Quirúrgica	<i>Fecha de intervención, profesional, número de pieza dental, protocolo de carga, exodoncia, expansión ósea, elevación de seno maxilar, regeneración de tejidos duros, regeneración de tejidos blandos, tiempo de colocación, tipo de hueso, indicación protésica, complicación quirúrgica.</i>
Seguimiento Postoperatorio	<i>Seguimiento postoperatorio, observación.</i>

Tabla 1: Dimensiones de "*BD_IMPLANTES.csv*".

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 del trabajo, no contienen errores, no se encuentran valores fuera de rango y no se han encontrado campos con valores 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 datasets final para iniciar con el diseño de un procedimiento que utilice

técnicas de minería de datos para resolver el problema planteado, 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 que representan al implante para obtener una nueva variable con datos del tratamiento se superficie, y así distintas composiciones (siempre teniendo en cuenta la opinión del experto) para ir experimentando y definiendo relaciones entre las variables. 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 de las variables ajustadas (eje x) en función de la variable label (SEGUI_POSTOP), es decir en base al éxito o fracaso del proceso de integración tejido - implante:

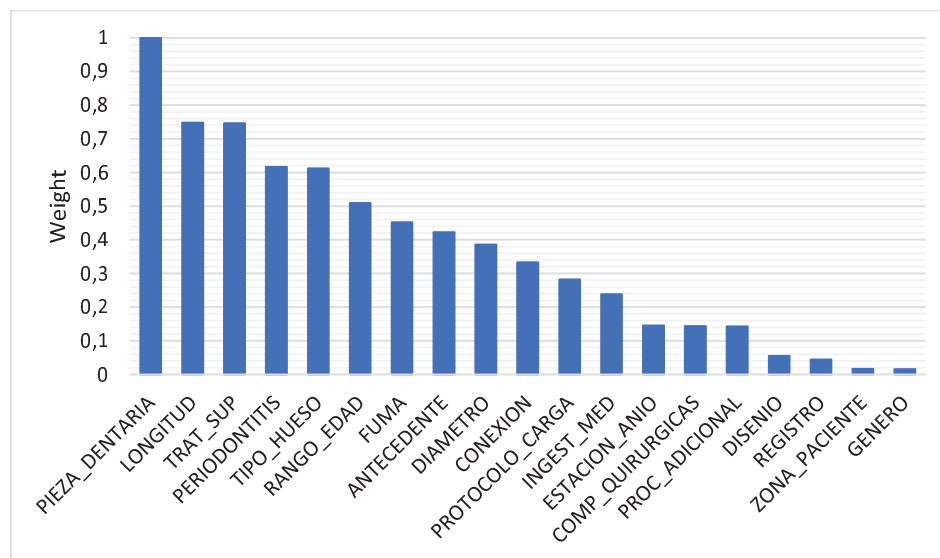


Figura 1: Ganancia de Información.

Actualmente, este trabajo se encuentra en la parte final de la tercera etapa de la metodología CRISP-DM, correspondiente a la preparación de los datos, para luego comenzar con la fase de modelado. Conjuntamente, se está profundizando el estudio y adaptabilidad del algoritmo máquina de vector soporte (SVM) y sus variantes, esto en función de los trabajos citados que utilizan y recomiendan la aplicación de este método para datasets de este tipo. Asimismo, se tiene en cuenta el estudio de lógica difusa, más específicamente la aplicación de prototipos deformables borrosos [6, 7].

REFERENCIAS

- [1] U. Fayyad, G. Piatetsky-Shapiro, P. Smyth, The KDD Process for Extracting Useful Knowledge from Volumes of Data, *Commun. ACM*, vol. 39, no. 11, 1996, 27–34.
- [2] J. Tamez, F. Zilli, L. Fandiño, J. Guizar, Factores relacionados con el éxito o el fracaso de los implantes dentales colocados en la especialidad de Prosthodoncia e Implantología en la Universidad de La Salle Bajío, *Revista Española de Cirugía Oral y Maxilofacial*, 2016, 1–9.
- [3] J. Domínguez, J. Acuña, M. Rojas, J. Bahamondes, S. Matus, Study of association between systemic diseases and dental implant failure, *Revista Clínica Periodoncia, Implantología y Rehabilitación Oral*, vol. 6, no. 1, 2013, 9–13.
- [4] A. Oliveira, C. Baldisserotto, J. Baldisserotto, A comparative study on machine learning techniques for prediction of success of dental implants, MICAI'05 Proceedings of the 4th Mexican international conference on Advances in Artificial Intelligence, 2005, 939–948.
- [5] P. Chapman, J. Clinton, R. Kerber, T. Khabaza, T. Reinartz, C. Shearer, R. Wirth, CRISP-DM 1.0: Step-by-step data mining guide, 2000, SPSS Inc.
- [6] J. Olivas Varela, From Fuzzy Deformable Prototypes to FuzzyWeb Search, *On Fuzziness*, vol. 298, 2013, 39–46.
- [7] R. M. Rodríguez, L. Martínez, V. Torra, Z. S. Xu, F. Herrera, Hesitant Fuzzy Sets: State of the Art and Future Directions, *Int. J. Intell. Syst.*, vol. 29, no. 2, 2014, 495–524.