

## Modelo de repositorio para gestionar reglas electorales generadas mediante técnicas de minería de datos

### Repository model to manage electoral rules generated through data mining techniques

LUNA-RAMÍREZ, Enrique<sup>1†\*</sup>, SORIA-CRUZ, Jorge, VELARDE-MARTÍNEZ, Apolinar y VILLALOBOS-ABARCA, Marco<sup>2</sup>

<sup>1</sup>Tecnológico Nacional de México, Instituto Tecnológico El Llano Aguascalientes, Km. 18 Carretera Ags.-S.L.P., C.P. 20230

<sup>2</sup>Universidad de Tarapacá, 18 de Septiembre 2222, Arica – Chile

ID 1<sup>er</sup> Autor: *Enrique, Luna-Ramírez* / ORC ID: 0000-0003-1818-7144, Researcher ID Thomson: S-8743-2018, CVU CONACYT ID: 122918

ID 1<sup>er</sup> Coautor: *Apolinar, Velarde-Martínez* / ORC ID: 0000-0002-6867-9414, Researcher ID Thomson: O-9756-2018, CVU CONACYT ID: 864001

ID 2<sup>do</sup> Coautor: *Jorge, Soria-Cruz* / ORC ID: 0000-0002-0616-1783, Researcher ID Thomson: T-1721-2018, CVU CONACYT ID: 103874

ID 3<sup>er</sup> Coautor: *Marco, Villalobos-Abarca* / ORC ID: 0000-0003-1843-0597, Researcher ID Thomson: W-1832-2019

DOI: 10.35429/JCA.2019.9.3.1.5

Recibido Enero 10, 2019; Aceptado Marzo 30, 2019

#### Resumen

Tener a la mano información sobre comportamientos electorales es de suma importancia para los diferentes actores que intervienen en un proceso electoral, ya que éstos, los comportamientos, juegan un papel mediático relevante durante las campañas políticas. En este sentido, en este artículo se propone un modelo de repositorio o meta-modelo para almacenar modelos y reglas electorales generadas mediante técnicas de minería de datos a partir de bases de datos que contienen resultados de procesos electorales recientes, particularmente, de las elecciones federales realizadas el año pasado en México para la elección presidencial y de las elecciones de este año en diferentes estados del país para elegir gobernadores, presidentes municipales y diputados locales. El meta-modelo propuesto incluye un conjunto de estructuras y una serie de procesos que permiten almacenar, con base en frases ad-hoc, los modelos y las reglas generadas, de manera que esta información pueda ser recuperada mediante alias de frases, emulando en cierta medida la ejecución de consultas en lenguaje natural. Así, este trabajo tiene una doble contribución, por una parte, el conocimiento electoral generado mediante minería de datos y, por otra, el meta-modelo para la gestión de este conocimiento.

**Meta-modelo, Reglas electorales, Minería de datos**

#### Abstract

Having information on electoral behaviors at hand is very important for the different actors involved in an electoral process, since these, the behaviors, play a relevant media role during political campaigns. In this sense, in this article, it is proposed a repository model or meta-model for storing models and electoral rules generated through data mining techniques from databases that contain results of recent electoral processes, particularly, of the federal elections held last year in Mexico for the presidential election and this year's election in different states of the country to elect governors, municipal presidents and local deputies. The proposed meta-model includes a set of structures and a series of processes that allow the models and rules generated to be stored, based on ad-hoc phrases, so that this information can be retrieved through phrase aliases, emulating in a certain way the execution of queries in natural language. Thus, this work has double contribution, on the one hand, the electoral knowledge generated through data mining and, on the other, the meta-model for the management of this knowledge.

**Meta-model, Electoral rules, Data mining**

**Citación:** LUNA-RAMÍREZ, Enrique, SORIA-CRUZ, Jorge, VELARDE-MARTÍNEZ, Apolinar y VILLALOBOS-ABARCA, Marco. Modelo de repositorio para gestionar reglas electorales generadas mediante técnicas de minería de datos. Revista de Cómputo Aplicado. 2019, 3-9: 1-5

\* Correspondencia del Autor (elunaram@hotmail.com)

† Investigador contribuyendo como primer autor.

## Introducción

En la actualidad, los procesos electorales a cualquier nivel, ya sea local o nacional, hacen uso de diversas tecnologías de la información para el seguimiento y control del gran cúmulo de datos que suelen generarse en este tipo de procesos (Soni *et al.*, 2017). No obstante, para un análisis más profundo de los datos generados, es decir, para extraer información oculta que no puede ser obtenida utilizando herramientas tradicionales para el análisis de información, se requieren herramientas especializadas y técnicas ad-hoc tales como las técnicas de minería de datos (Shejale *et al.*, 2016). De esta manera, es posible extraer conocimiento valioso de los resultados de un proceso electoral, como podría ser el comportamiento de los votantes bajo ciertas circunstancias, entre otras variables.

Con base en esto, en este artículo se propone un modelo de repositorio para gestionar reglas y modelos electorales, generados a partir de bases de datos de diversos procesos electorales con la ayuda de técnicas de minería de datos. El artículo ha sido estructurado en seis secciones, iniciando con la presente introducción y continuando en la sección dos con la descripción de la metodología utilizada para desarrollar nuestro trabajo. En la sección tres, se discute el estado del arte, y en la sección cuatro, se describe el modelo objeto de este trabajo. En la sección cinco, se presentan los resultados obtenidos a partir de un prototipo construido con base en el modelo propuesto. Finalmente, en las secciones seis y siete, se dan las conclusiones sobre el trabajo realizado y las referencias correspondientes, respectivamente.

## Metodología

En principio, para el desarrollo de este trabajo, se generaron diversos modelos a partir de bases de datos relativas a procesos electorales, para lo cual se siguió la metodología descrita por Luna-Ramírez *et al.* (2018), misma que puede ser resumida en la ejecución de las etapas de un proceso típico de minería de datos, que inicia con la integración y preprocesamiento de datos, y culmina con la generación de conocimiento, que en este caso es generación de modelos y reglas.

Una vez que se contó con un banco de modelos y reglas electorales (en formato de imagen), se procedió al diseño del modelo de repositorio para la gestión de estos elementos, para lo cual se definió en primera instancia su estructura genérica, es decir, sus componentes principales.

Posteriormente, se definieron las estructuras y procesos necesarios para llevar a cabo la gestión de modelos y reglas, incluidas las estructuras que permiten recuperar información del repositorio mediante consultas en lenguaje natural.

Como última etapa, se construyó un prototipo basado en el diseño final del repositorio, mismo que fue sometido a pruebas de funcionalidad por parte de usuarios técnicos, quedando pendiente las pruebas de usabilidad por parte de usuarios finales como podrían ser consejeros de institutos electorales, representantes de partidos políticos, casas encuestadoras, o personas interesadas en el tema electoral.

## Estado del arte

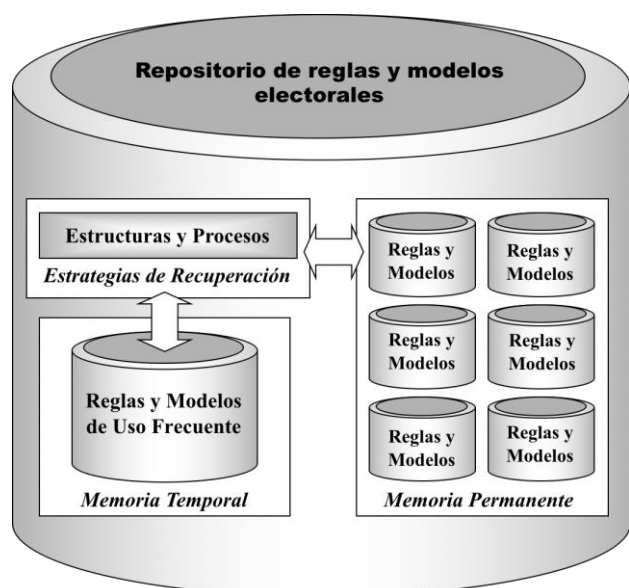
Galbrun & Miettinen (2016) proponen utilizar *redescription mining* (minería de redescipción) para encontrar de manera automática posturas políticas de gente de un cierto nicho socio-económico, así como nichos relacionados entre sí. Aseveran que esta técnica puede ser aplicada sobre información disponible en aplicaciones de seguimiento electoral, ello con la finalidad de proveer pronósticos sobre el posicionamiento de los candidatos en un proceso electoral.

Poloni & Formolo (2015) proponen un sistema para mejorar la seguridad en una elección que se realiza de manera electrónica, basado en el análisis del comportamiento de los votantes para detectar resultados disonantes. El sistema utiliza técnicas de minería de datos para detectar eventos sospechosos, con buenos resultados en la detección de fraudes de acuerdo a los autores. Sangar *et al.* (2013) propusieron un modelo de para predecir la inclinación de los votantes en la onceava elección presidencial de la República Islámica de Irán. Su modelo incluyó las técnicas de minería de datos de *árboles de clasificación*, *Naïve Bayes* y *KNN*. Para probar su modelo, los autores seleccionaron a personas que asistieron a dicha elección y anticiparon su decisión.

Wegrzyn-Wolska & Bougueroua (2012) propusieron un sistema para analizar las tendencias en la elección presidencial francesa a partir de los comentarios en Twitter utilizando *minería de texto*. Según los autores, el sistema realiza de manera automática la colección, evaluación y clasificación de los tweets para analizar las tendencias.

### Modelo propuesto

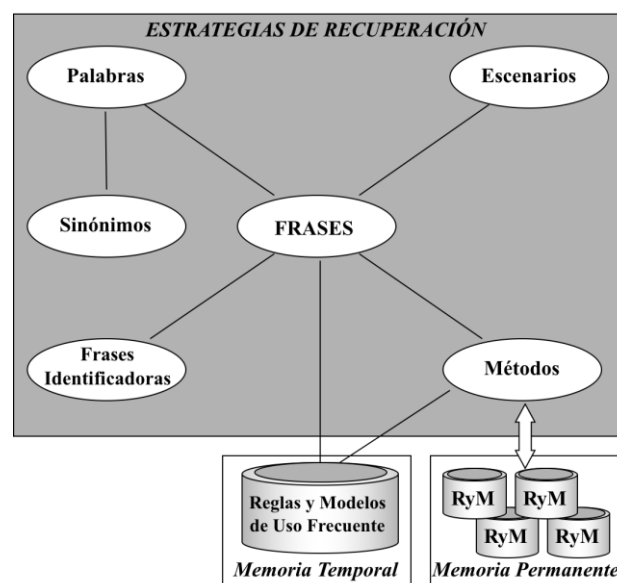
En la figura 1, se presenta nuestra propuesta de modelo de repositorio para gestionar reglas y modelos electorales. El repositorio consiste en tres componentes: una memoria permanente para almacenar todo el conocimiento extraído de bases de datos electorales (reglas y modelos), una memoria temporal para almacenar datos de uso frecuente y un conjunto de estrategias que permiten la interacción entre ambas memorias para la recuperación de datos.



**Figura 1** Modelo de repositorio para almacenar reglas y modelos electorales

Es importante señalar que, bajo este enfoque, cuando se realiza una consulta al repositorio, las estrategias de recuperación operan primero sobre la memoria temporal, que se supone significativamente más pequeña que la memoria permanente, por lo que la consulta podría ser inmediata. No obstante, si la información buscada no se localiza en la memoria temporal, se acude a la memoria permanente para extraerla y después almacenarla en la memoria temporal, desde donde se dará respuesta a la consulta.

En el caso de las estrategias de recuperación, éstas consisten en estructuras y procesos con una interdependencia funcional, debido a que las estructuras almacenan la información que los procesos necesitan para dar respuesta a las consultas del usuario final. En la figura 2, se muestran las estructuras consideradas en nuestro modelo de repositorio. Como se puede observar en la figura, la estructura FRASES es la parte central de las estrategias de recuperación, cuya función es almacenar todas las claves que interrelacionan las diferentes estructuras.



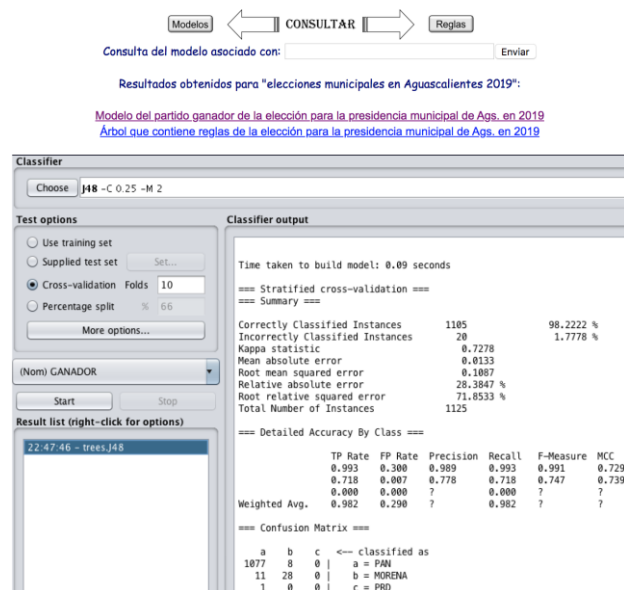
**Figura 2** Estrategias para recuperar las reglas y modelos almacenados en el modelo de repositorio

Así, para recuperar una regla o un modelo de la memoria permanente y almacenarlo(a) en la memoria temporal, en la estructura FRASES se “dispara” el método correspondiente a la consulta realizada por el usuario, una vez que ésta haya sido tratada de la siguiente manera: la frase (consulta) ingresada por el usuario es descompuesta en las palabras que contiene, para posteriormente determinar los sinónimos de cada palabra con el propósito de generar diversas combinaciones de palabras, alguna de las cuales coincidirá con la frase que identifica al modelo o regla que se desea recuperar. Es importante señalar que como parte de nuestra propuesta, se consideran escenarios asociados a consultas, es decir, la respuesta a una determinada consulta, típicamente será, además del modelo o regla que se desea, un conjunto de modelos o reglas asociados(as) al tópico de la consulta, de manera que la respuesta proporcionada por el repositorio sea lo más completa posible.

## Resultados

Con base en el modelo de repositorio propuesto en la sección anterior, se generó un prototipo que permite almacenar reglas y modelos electorales, y realizar consultas de la manera antes descrita. A manera de ejemplo, en la figura 3, se muestra un modelo relacionado con los resultados de las elecciones municipales en Aguascalientes en el presente año, el cual fue generado con la ayuda de la herramienta Weka (2019), y almacenado en el repositorio en formato de imagen.

### - Repositorio de Modelos y Reglas Electorales -

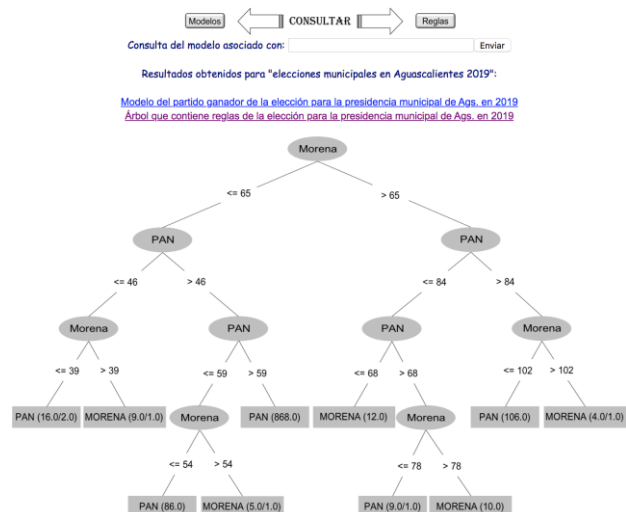


**Figura 3** Modelo de reglas electorales generado mediante Weka, almacenado en el modelo de repositorio propuesto

Como se puede observar en la figura, para una consulta (en lenguaje natural) relacionada con el tema antes mencionado, se obtuvo un escenario compuesto por dos enlaces correspondientes a dos representaciones diferentes de un modelo que describe información interesante sobre las pasadas elecciones en Aguascalientes (patrones, tendencias, etc.).

La figura anterior corresponde precisamente a la primera representación, en la cual se puede apreciar la técnica de minería de datos utilizada para generar el modelo (el clasificador *J48*), así como el algoritmo utilizado para validar el modelo (*cross-validation*). En la figura 4, se muestra la otra representación del modelo, en forma de árbol.

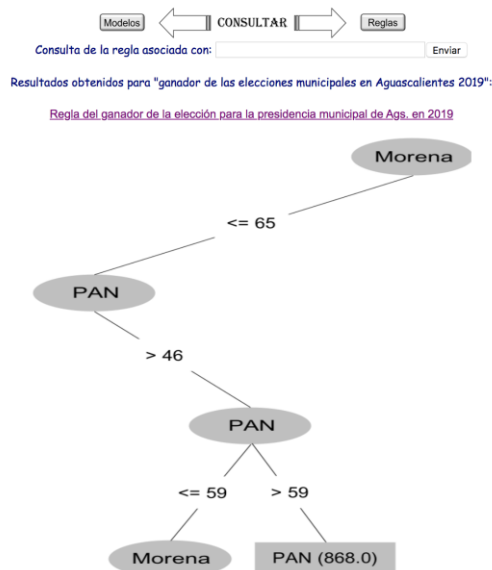
### - Repositorio de Modelos y Reglas Electorales -



**Figura 4** Modelo de reglas electorales en representación de árbol, almacenado en el modelo de repositorio propuesto

La representación de árbol es particularmente útil, ya que de ella se pueden obtener de manera inmediata reglas (patrones) de comportamiento de los votantes. Como ejemplo de ello, véase la figura 5, la cual se muestra la regla más relevante (contenida en el modelo de árbol), asociada al tema que se está ejemplificando, las elecciones de Aguascalientes municipio en el presente año.

### - Repositorio de Modelos y Reglas Electorales -



**Figura 5** Ejemplo de una regla electoral extraída del modelo de árbol, almacenada en el modelo de repositorio propuesto

En esta regla, se puede apreciar con claridad el comportamiento de los votantes en 868 de las 1125 casillas que operaron en el municipio de Aguascalientes: si el número de votos a favor de MORENA es menor o igual que 65 y el número de votos a favor del PAN es mayor que 59, con toda seguridad la victoria será de este último partido, independientemente de cualquier otra circunstancia. Esta regla se cumplió en el 77% de las casillas (868 de 1125), lo cual podría ser de significancia en la definición de estrategias para procesos electorales futuros.

### Conclusiones

En principio, resaltar la doble contribución de nuestro trabajo, ya que por una parte, se generó conocimiento electoral mediante la aplicación de técnicas de minería de datos y, por otra, se diseñó un modelo de repositorio para gestionar dicho conocimiento, de forma tal que éste pueda estar disponible de manera inmediata a los usuarios. Así, el repositorio y la información contenida en él, podría servir de apoyo en la definición de estrategias para futuros procesos electorales.

Se describió la forma como opera el sistema de repositorio con base en ejemplos específicos referentes al proceso electoral realizado en el presente año en Aguascalientes; no obstante, nuestro modelo de repositorio está diseñado para operar con información generada a partir de cualquier tipo de proceso electoral.

### Referencias

Galbrun, E. & Miettinen, P. (2016). "Analysing Political Opinions Using Redescription Mining". *Proceedings of the IEEE 16th International Conference on Data Mining Workshops*, pp. 422-427.

Luna-Ramírez, E., Soria-Cruz, J., Cruz-Valenzuela, R. y Castillo-Zúñiga, I. (2018). "Análisis de los Resultados en las Elecciones Federales Mexicanas 2018 mediante técnicas de minería de datos". *Revista de Análisis Cuantitativo y Estadístico* (ISSN- 2410-3438), Vol.5, No.16, pp.6-10.

Poloni, Y. T. & Formolo, D. (2015). "Data Mining to Identify Fraud Suspected on Electronic Elections". *Proceedings of the 9th International Conf. on Complex, Intelligent, and Software Intensive Systems*, pp. 19-23.

Sangar, A. B., Khaze, S. R. & Ebrahimi, L. (2013). "Participation Anticipating In Elections Using Data Mining Methods". *International Journal on Cybernetics & Informatics (IJCI)* Vol.2, No.2, pp. 47-60.

Shejale, A. & Gnagawane, V. (2016). "An implementation of efficient techniques for tree based mining in human social dynamics". *Proc. of the Intnl. Conference on Data Mining and Advanced Computing (SAPIENCE)*, pp: 8-16.

Soni, D., Sharma, M. & Khatri, S. K. (2017). "Political opinion mining using E-social network data" *Proc. of the International Conference on Infocom Technologies and Unmanned Systems (ICTUS)*, pp. 163-165.

Wegrzyn-Wolska, K. & Bougueroua, L. (2012). "Tweets mining for French Presidential Election". *Proceedings of the 4th International Conference on Computational Aspects of Social Networks (CASoN)*, pp. 138-143.

Weka: Machine Learning Software in Java. Recuperado de <https://www.cs.waikato.ac.nz/~ml/weka/index.html>, 12 de agosto de 2019.