



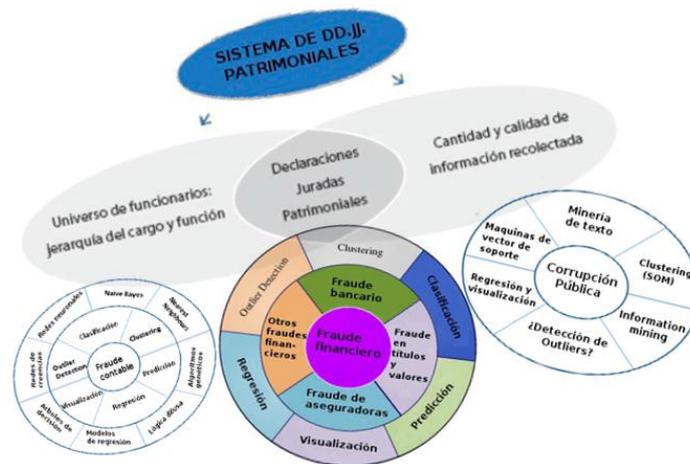
Editorial de la
Universidad Tecnológica Nacional
U.T.N. – Argentina

Trabajo final Integrador

Propuesta de detección de datos anómalos y ruido para un caso de contralor civil

Lic. Rodrigo López-Pablos

Director: Prof. Dr. Horacio Daniel Kuna



Especialización en Ingeniería en Sistemas de la Información

Escuela de Posgrado

Facultad Regional Buenos Aires

Universidad Tecnológica Nacional – U.T.N.

Junio de 2016

Editorial de la Universidad Tecnológica Nacional - edUTecNe

<http://www.edutecne.utn.edu.ar>

edutecne@utn.edu.ar

©[Copyright]

edUTecNe, la Editorial de la U.T.N., recuerda que las obras publicadas en su sitio web son de libre acceso para fines académicos y como un medio de difundir la producción cultural y el conocimiento generados por autores universitarios o auspiciados por las universidades, pero que estos y edUTecNe se reservan el derecho de autoría a todos los fines que correspondan.

ÍNDICE

pág.

1. Introducción.....	7
1.1 Objetivos del TFI.....	9
1.2 Alcance.....	10
1.3 Contenido del TFI.....	10
2. El problema de la corrupción pública.....	12
2.1 La corrupción, su naturaleza compleja y su impacto societal.....	12
2.2 El problema del acceso ciudadano a la información pública.....	15
3. Estado de la cuestión	17
3.1 La minería de datos.....	17
3.2 Minería de datos para la detección de fraude y corrupción.....	21
3.3 Métodos para la detección de outliers.....	26
4. Solución para un caso de contralor.....	29
4.1 Medidas para la prevención de la corrupción.....	30
4.1.2 Los sistemas de DDJJ como mecanismo de prevención de la corrupción.....	32
4.2 Propuesta de aproximación metodológica para la detección de ruido y anomalía en DDJJ.....	35
4.2.1 Propuesta de procedimientos para la detección de outliers y ruido en BBDD.....	35
4.2.2 Metodología híbrida para la detección de outliers.....	37
4.2.3 Propuesta de materiales y datos para la experimentación.....	38
5. Conclusiones.....	39
Bibliografía.....	41

ÍNDICE DE TABLAS

pág.

1. Relación entre técnicas y métodos de minería de datos.....	20
2. Enfoques para el problema de la detección de outliers.....	28
3. Aspectos de los gobiernos abiertos.....	32
4. Algoritmos por procedimiento híbrido para la detección de outliers y ruido.....	36
5. Procedimientos según entorno, algoritmos y enfoque.....	37

ÍNDICE DE FIGURAS

pág.

1. Contextos y escenarios de corrupción	13
2. Impacto de la corrupción.....	14
3. Proceso KDD.....	17
4. Clasificación de técnicas de explotación de la información para la detección de fraude financiero.....	23
5. Algoritmos y técnicas de minería de datos para el descubrimiento de fraude fiscal.....	24
6. Técnicas de EI usados para la detección de corrupción pública	25
7. Componentes de las técnicas de detección de anomalías.....	27
8. Sistema de DDJJ patrimoniales.....	33

NOMENCLATURAS

BD/BBDD	Base de datos/Bases de datos.
CAATs	Computer Assisted Audit Techniques. Técnicas de Auditoría Asistidas por Computador.
C4.5	Algoritmo de inducción desarrollado por Quinlan (1993).
DBSCAN	Algoritmo de detección de datos anómalos basado en densidad.
DJ/DDJJ	Declaración Jurada/Declaraciones juradas.
EI	Explotación de la información.
K-Means	Algoritmo que permite clasificar un conjunto de objetos en un número K de clusters.
LOF	Local Outlier Factor. Factor de anomalía local.
LRD	Local Reachability Density. Densidad de accesibilidad local.
OA	Oficina anticorrupción.
OGP	Open Government Partnership. Alianza para el gobierno abierto.
PRISM	Algoritmo de aprendizaje basado en reglas.
SOM	Self-Organizing Map, Mapas auto-organizados.
TDIDT	Top Down Induction Decision Trees. Árboles de decisión de arriba hacia abajo.
TI	Teoría de la información.

1. INTRODUCCIÓN

La ingeniería de explotación de la información y la minería de datos han sido aplicadas generalmente a cuestiones dentro del ámbito privado, de los negocios y las corporaciones en general, no tanto así como un conjunto de técnicas al servicio de los organismos del Estado ni para el ejercicio del contralor Ciudadano, al menos en el caso Argentino, no se observa un desarrollo sustantivo de esta tecnología en administración pública.

Sus campos de aplicación interdisciplinar abordaron generalmente la utilización de estas técnicas como a herramientas vinculadas a las problemáticas frecuentes ligados a la empresa privada o dirigidas a la gran corporación lo que cuadró la morfología semántica del campo del arte; consecuencia de ello, se observa en su uso aplicativo donde se suele referirse a “negocios” y “clientes” en lugar de “ciudadanos” o “contribuyentes”, así como a “consumidores” en lugar de usuarios de servicio público, servicio comunitarios, etc.

En este esfuerzo se propone abrir los indicios que las herramientas que ofrecen las técnicas de EI en un área donde han sido sub-explotadas tanto por la administración pública como por los organismos de contralor, por ende presentando un vacío epistémico. Estas técnicas tampoco han sido consideradas en casos específicos de casos para el control cívico Ciudadano a escala individual, transparencia administrativa o resorte de proyección de organizaciones civiles encargadas del control y la auditoría cívico-ciudadana de las personas.

Sin una sociedad civil robusta peligran los derechos de los individuos y de los grupos que no adhieren incondicionalmente al sistema, el fortalecimiento de las sociedades civiles requiere como condición de posibilidad, la potenciación de una ética compartida por todos los miembros de esa misma sociedad (Cortina, 1994). Al tiempo presente, tal esfuerzo vuelve necesaria el uso de procesos informáticos como herramental mínimo para asegurar su fortalecimiento.

Siendo que la información es la columna vertebral y epicentro organizativo para la toma de decisiones organizacional de cualquier entidad tanto pública como privada, la calidad de los datos, y el conocimiento que de allí se explote guarda una importancia sin precedentes para el tratamiento de los datos en las instituciones.

Los estudios e investigaciones científicas del fenómeno de la corrupción, a través del uso de técnicas de minería de datos, estuvieron enfocados generalmente a la apreciación de la corrupción desde niveles de percepción social, así como de relacionar supuestas conectividades entre variables macroeconómicas y percepción de la corrupción pero no directamente sobre los sujetos de poder que de caer en actos corruptivos más perjuicio desencadenaría en el tejido social, i. e. sobre los representantes sociales que deben garantizar el bienestar general y el cumplimiento de la constitución, leyes y normas de convivencia.

La *transparencia* depende conceptualmente de la calidad de la información organizacional que se audita, ésta misma esta definida como la *calidad de un gobierno, empresa, organización o persona de abierta a la divulgación de información, normas, planes, procesos y acciones* (International Transparency, 2009). En tiempos digitales, la continuación natural de políticas de transparencia se traduce en accesos a la información pública más efectivos y eficientes para la ciudadanía, así como la capacidad de brindar herramientas efectivas y la capacitación necesaria al auditor Ciudadano para el control de sus propios representantes.

En términos republicanos, la democracia indirecta representativa depende, cada vez más e inevitablemente, de Ciudadanos informados capaces de ejercer sus derechos civiles, y tener representantes probos que como paradigma Ciudadano obren con el ejemplo pues encarnan durante un tiempo el poder delegado de los habitantes. Dada la complejidad de inherente a toda realidad social las herramientas de EI se vuelven una herramienta más de la ciudadanía y los Estados para su auto control y regulación.

Mediante esta propuesta se introduce la utilización de técnicas de minería de datos en el ámbito público en general y contribuir en los procesos cívicos y de control en particular.

En este sentido estas nuevas técnicas contribuirán a hacer operativo el derecho a la información pública lo cual implica en términos de prevención de la corrupción, dar a los Ciudadanos las herramientas para ejercer sus derechos y controlar el accionar del Estado (Naciones Unidas, 2004; Raigordsky y Geler, 2007) brindando las herramientas computacionales para crear conocimiento a partir de esa información se estará contribuyendo al empoderamiento de esa misma ciudadanía.

Nunca antes se había propuesto la utilización de datos provenientes de los mismos entes de auditoría y control del propio Estado para potencial allí, a través de los mecanismos institucionales constituidos para esa tarea, para explotar información y generar conocimiento a través del uso de estas técnicas.

Por otra parte, también es la primera vez que se propone la utilización de técnicas específicas de detección de outliers para ayudar a acometer dichos objetivos cívico institucionales.

1.1. OBJETIVOS DEL TFI

En vista a los problemas vinculados que aquejan a la sociedad contemporánea, se propone la posibilidad de utilización de un modelo de procesos para la utilización de procedimientos de explotación de la información desarrollado por Kuna (2014); que en su aplicación combinada, permita la detección de valores anómalos y ruido en bases de datos en organismos públicos de cualquier índole. Por otra parte, como propuesta empírica hipotética no aplicada en este trabajo; se sugiere la experimentación de BBDD públicos, como lo son las declaraciones juradas obligatorias de funcionarios públicos para el caso Argentino (DDJJ).

Estos pueden enumerarse de la siguiente manera:

1. Proponer la utilización de técnicas y herramientas de detección de datos anómalos y explotación de la información para el descubrimiento y revelación de comportamiento corrupto como parte de la función cívica ciudadana.

2. Proponer y considerar las técnicas de explotación de información sobre BD de datos públicos (en este caso BD de DDJJ) como materia prima e instrumento para la lucha contra la corrupción.
3. Contribuir a la formación del Ciudadano auditor que haga al control cívico de la información Estatal.

1.2. ALCANCE

El alcance extendido de este trabajo integrador comprende la propuesta de aplicación de procesos de explotación de la información en el ámbito público, ya sea para los auditores dentro de la administración pública como para el ejercicio cívico y de contralor de la propia ciudadanía representada; *i.e.*, lo que hace a las funciones de las organizaciones civiles.

En base a su aplicación efectiva en datos públicos reales se justifica la necesidad de acceso a la ciudadanía de datos públicos de manera de contribuir a la sociedad en su capacidad de auto-gestión y auto-auditoría a través de organizaciones civiles y del propio Estado contralor.

1.3. CONTENIDO DEL TFI

Siendo el presente Capítulo el correspondiente a la introducción del TFI. A continuación su contenido y estructura básica.

Capítulo 2: Se plantea el problema de la corrupción pública y las hipótesis de aplicación de procesos de explotación de la información para ayudar al combate de su flagelo.

Capítulo 3: Se describe el estado del arte sobre la auditoría de sistemas en organismos públicos y privados, así como los antecedentes en la utilización de técnicas de minería de datos para la detección de fraude.

Capítulo 4: Se desarrolla la solución propuesta para un caso de contralor civil proponiendo el uso de DDJJ como eje para dicha tarea, y planteando la descripción de distintas técnicas, procesos y metodologías a ser empleadas para la detección de datos anómalos.

Capítulo 5: Se exponen las conclusiones finales del TFI.

Sujeto a esta estructura, inmediatamente a continuación se presentan los problemas e hipótesis correspondientes.

2. EL PROBLEMA DE LA CORRUPCIÓN PÚBLICA

En este Capítulo se presenta la problemática general de la corrupción en las sociedades, su naturaleza, taxonomía y las políticas que posibilitan la lucha al flagelo. Se describe su naturaleza e impacto así como la importancia central del acceso a la información pública.

2.1. LA CORRUPCIÓN, SU NATURALEZA COMPLEJA, Y SU IMPACTO SOCIETAL

La corrupción es un flagelo que vulnera los derechos humanos de las personas, acentúa la desigualdad social y afecta el desarrollo de la población. El sistema democrático requiere, necesariamente, del aporte de la ciudadanía en el fortalecimiento de las instituciones, como un ejercicio indirecto de la soberanía del pueblo representado.

El 9 de diciembre de 2013, el secretario general de la ONU Ban Ki-Moon, se dirigía a la asamblea general con las siguientes palabras¹:

“La corrupción impide el crecimiento económico al elevar los costos y socava la gestión sostenible del medio ambiente y los recursos naturales. Así mismo, quebranta los derechos humanos fundamentales, agrava la pobreza e incrementa la desigualdad al desviar fondos de la atención de la salud, la educación y otros servicios esenciales. Los efectos perniciosos de la corrupción los sienten miles de millones de personas en todo el mundo” (UNODC, 2013)

En sentido amplio la corrupción se entiende como “el abuso del poder para beneficio propio” (Transparency International, 2009) el abuso autoritario del poder, el cual algunos autores puede comprender:

¹Un año después a este pronunciamiento, el 9 de diciembre de 2014 se declaraba esa misma fecha como día mundial contra la corrupción.

- El mal uso del poder político.
- Un poder encomendado que puede estar tanto en el sector privado como en el público.
- Un beneficio particular, referido a beneficios personales para la persona que hace mal uso del poder, incluyendo también a miembros de su familia inmediata y de sus amigos.

La definición por otra parte de corrupción política es entendida también como la manipulación de políticas, instituciones y normas de procedimiento en la asignación de recursos y financiamiento por parte de los responsables de las decisiones políticas, quienes se abusan de su posición para conservar su poder, estatus y patrimonio (Transparency International, 2009).

Por otra parte el hecho corruptivo puede tasarse de diferentes maneras: (i), como *corrupción negra*, a aquellas acciones ocultas como el soborno y la extorsión; (ii), *corrupción blanca*, cuando el acto corruptivo es aceptado ampliamente por la comunidad y la *corrupción gris* en donde existe un debate abierto en la sociedad y medios sobre si esta se considera correcta o no (Cisneros, 2011; Heidenheimer et. al., 1989).

Desde la complejidad del fenómeno de la corrupción, la percepción y aceptación de los hechos de corrupción en las sociedades puede asociarse a contextos y escenarios particulares; por ejemplo, una sociedad puede encontrarse en algún punto medio entre dos escenarios o contextos corruptivos: una sociedad completamente corrupta donde la misma es estructural o una donde se ve mínimamente afectada donde la corrupción se presenta de forma marginal (Gómez y Bello, 2009). Los contextos y escenarios donde el fenómeno corruptivo marginal o estructural puede variar como presenta la Figura 1.



Figura 1: Contextos y escenarios de corrupción (Gómez y Bello, 2009).

Los mismos autores describen las consecuencias de una sociedad que sufre un escenario de corrupción estructural la cual se caracteriza por las siguientes:

- Pérdida de legitimidad del sistema político.
- Asignación de recursos ineficiente e ineficaz.
- Destrucción del profesionalismo y las capacidades de las personas.
- Segregación, marginación y desánimo de las personas honestas.
- Pérdida de previsibilidad sobre el futuro de las organizaciones del sistema.

Por otra parte, el impacto político y social de la corrupción en la sociedad es múltiple y por su particularidad compleja es difícil de mensurar directamente. Para comprender el verdadero alcance del mismo es necesario entender el fenómeno de manera integral, en donde el perjuicio social del acto corruptivo se encuentra mediado por la concurrencia de otros hechos y situaciones que impacta colateralmente en múltiples víctimas indirectas, impactando de lleno en lo público y acrecentando las diferencias sociales, como se observa en la Figura 2.



Figura 2: Impacto de la corrupción (Gómez y Bello, 2009).

Como se aprecia en la Figura 2, la corrupción no solo afecta la estabilidad política y mina la confianza societaria -la cual se percibe por ejemplo al ver un noticiero- esta también tiene consecuencias indirectas en la desigualdad y el empobrecimiento

poblacional, y consecuentemente sobre el capital social. Sin embargo, cuando el impacto del incremento y presencia de la corrupción se ve amplificado en el empeoramiento del escenario en la corruptela social, la percepción de su perjuicio en la misma -si la sociedad cuenta con los mecanismos cívicos e institucionales para canalizar su curso- se traduce en un mayor participación y control cívico de la propia población como auto-reflejo de esa percepción, cuando esta percepción es genuina.

De lo contrario, si la propia sociedad víctima y culpable al mismo tiempo de un escenario corruptivo no cuenta con las herramientas institucionales, técnicas y ciudadanas provistas por el propio Estado, las sociedades civiles y la ciudadanía se encontrarán con mayores inconvenientes y una disipación de sus energías para lograr un Estado más efectivo en el logro del bienestar.

2.2. EL PROBLEMA DEL ACCESO CIUDADANO A LA INFORMACIÓN PÚBLICA

El acceso a la información pública es uno de los cuatro pilares fundamentales necesarios, junto a la transparencia fiscal, la revelación patrimonial de los funcionarios oficiales y el compromiso Ciudadano, para la existencia de un gobierno abierto multiparticipativo (OGP, 2012).

En Argentina aún no existe ni se ha planteado una estrategia en como sería una mejor participación digital ciudadana (Vercelli, 2013), si bien existe un proyecto de ley de acceso a la información pública este aún no ha sido promulgado como tal lo que hace de factor negativo adicional, además de la necesaria formación civil del Ciudadano, la no disponibilidad de la información pública de libre acceso al Ciudadano, coarta las posibilidades de una vida democrática plena a la ciudadanía para una participación activa y responsable de todos los actores de la sociedad en la construcción de los consensos necesarios para el fortalecimiento de todo sistema social cívico.

En este sentido el acceso de la ciudadanía a la información pública guarda un rol central en la acción cívica de la misma cuando se afirma que *“Un Ciudadano que enfrenta vacíos y lagunas de información sobre los asuntos públicos no tendrá la oportunidad de*

expresar su opinión en temas inherentes a la administración gubernamental, por lo que mucho menos podrá juzgar los actos de sus gobernantes y como consecuencia a la hora de ejercer su derecho de elegir a estos poca certeza tendrá de que su elección sea correcta; quebrantándose de esa forma uno de los pilares fundamentales, la participación ciudadana, que sustenta a todo el Estado de derecho.” (Cisneros, 2011). El acceso a la información pública es fundamental para que desde la propia base poblacional ciudadana permita el auto-control y auditoría cívica de los gobernados, y que, naturalmente para la naturaleza digital de las transacciones informacionales, es ineludible que los datos a los que aquellos accedan sean de la mejor calidad posible.

Si bien el recabo de información se encuentra oficiado por la Oficina Anticorrupción (OA²), el acceso a las mismas se hace por pedido particular o colectivo lo que dificulta el acceso rápido y efectivo a los datos. Además de problemas burocráticos de modernización en el dominio público, como lo son el necesario y obligado traspaso de archivos desde el formato papel al digital; la OA, se enfrenta a dificultades de excesiva centralización, lo que resulta inconveniente para realizar controles formales y sustantivos, obstáculos en el acceso público a las declaraciones juradas de funcionarios. (Ferro, Giapponi y Gómez, 2007). A pesar de que tales DDJJ presentan algunos errores, la falta de difusión de los mismos descubre la carencia de una estrategia Estatal frente a sistemas cívico-institucionales e informacionales frente a una sociedad digitalmente más demandante.

La disposición pública de esa información pone de relieve, además de la impostergable formación digital de la ciudadanía, la necesidad de contar con herramientas para el análisis y explotación de la información. Si bien se han hecho intentos desde las ciencias de la computación para encontrar soluciones tecnológicas para la mejora de los procesos de participación ciudadana directa (Colombo et al., 2013); pocos han sido los casos que encaren la participación indirecta que ejerce el individuo como auditor y contralor cívico de sus propios representantes así como de sociedades civiles abocadas a esa tarea. Por esos motivos, en el siguiente Capítulo, se repasan los antecedentes en el campo para conocer los verdaderos alcances de una aplicación herramental.

² Al tiempo de la confección de la totalidad de este trabajo integrador en toda su extensión, la Oficina Anticorrupción pasó a llamarse Secretaría de Ética Pública, Transparencia y Lucha contra la Corrupción.

3. ESTADO DEL CUESTIÓN

En este Capítulo se expone el estado de la cuestión en el campo de la minería de datos y los procesos de explotación de la información para la detección de datos anómalos, en la sección subsiguiente se repasa el estado del arte sobre minería de datos así como se describen los antecedentes correspondientes a la utilización de dichas técnicas avocadas al descubrimiento de fraude y corrupción pública y privada. Posteriormente también se abordan los métodos y enfoques para la detección de datos anómalos.

3.1. LA MINERÍA DE DATOS

La minería de datos puede definirse como un proceso de extracción no trivial del conocimiento disponible de manera explícita en BBDD (Schiefer et al., 2004; Clark, 2000); siendo que tal conocimiento en cuestión previamente desconocido, resulta posteriormente adquirido de alguna utilidad inherente para la sociedad en general o en particular.

Por esos motivos la minería de datos puede ser enmarcada como un elemento dentro de un proceso más general dirigido al descubrimiento de conocimiento dentro de grandes sets de datos “*Knowledge Discovery in Databases*” (KDD) (Fayyad et al., 1996); figurativamente, aquel representa la tercera fase en el proceso de descubrimiento de conocimiento como se presenta en el siguiente cuadro.

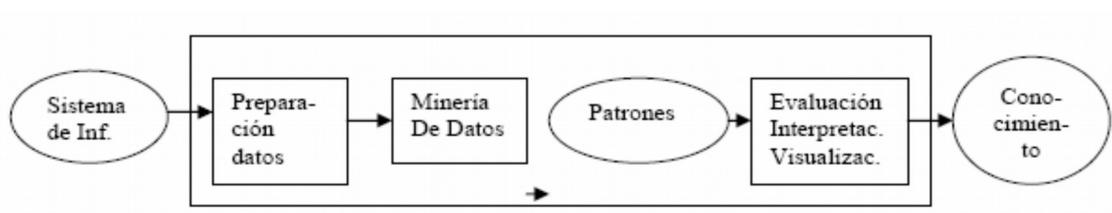
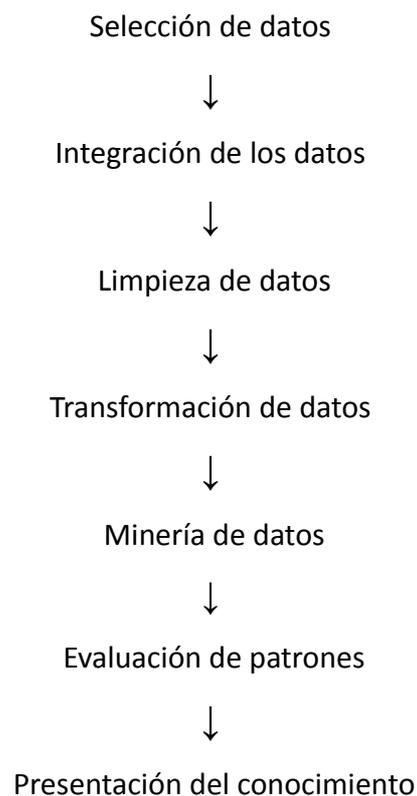


Figura 3: Proceso KDD (Kuna, 2014).

Como se desprende de la Figura 3, la minería de datos es un componente más de un proceso más general de descubrimiento de conocimiento en BBDD, precedido por la etapa de preparación de los datos y proseguido por la de obtención de patrones, producto de la aplicación de técnicas de minería de datos y materia prima para la

obtención de conocimiento *a posteriori* de la fase de evaluación e interpretación de patrones.

Análogamente, la minería de datos también puede ser considerada parte de un proceso de EI (Larose, 2005), dentro del cual la minería de datos viene a conformar la quinta fase del proceso (Britos et al., 2005): *a priori* de la evaluación de patrones y *a posteriori* de la transformación de datos como se describe a continuación.



Modelos y técnicas de la minería de datos

A su vez los modelos de minería de datos pueden ser tipificados de dos maneras básicas:

- Modelos predictivos: Se utiliza fundamentalmente la clasificación donde según una BD, donde se busca predecir a qué clase pertenece una nueva instancia, asignándose previamente un valor en cada tupla correspondiente a una clase predeterminada. Para análisis predictivos numéricos se utiliza la regresión.

- Modelos descriptivos: Buscando generar etiquetas y/o agrupaciones estos modelos exploran las propiedades de los datos bajo análisis, donde se utiliza fundamentalmente el clustering así como la correlación y factorizaciones para evaluar el grado de similitud entre atributos numéricos; reglas de asociación, para la búsqueda de asociaciones inter-atributos no explícitas.

Las técnicas de minería de datos por otra parte pueden clasificarse entre técnicas basadas en análisis estadístico y técnicas basadas en sistemas inteligentes. Por un lado algunas de las tecnologías basadas en análisis estadístico más usadas son las siguientes.

- Análisis de agrupamiento
- Análisis de varianza
- Análisis discriminante
- Prueba Chi-cuadrado
- Regresión
- Series de tiempo

Por otra lado, algunas de las tecnologías basadas en sistemas inteligentes pueden enumerarse de la siguiente manera (García Martínez et al., 2003).

- Algoritmos de inducción
- Perceptrón multicapa
- Máquina de vector soporte
- Algoritmos *a priori*
- Redes neuronales SOM
- Algoritmos genéticos
- Redes bayesianas
- Algoritmo del vecino más próximo
- Variedad de técnicas estocásticas, relacionales, declarativas, difusas, así como múltiples tecnologías híbridas.

Resumidamente, puede explicitarse relacionalmente los vínculos entre técnicas y modelos de minería de datos en la siguiente Tabla 1.

Técnicas	Predictivo		Descriptivo		
	Clasificación	Regresión	Agrupamiento	Regla de Asociación	Correlaciones/ Factorizaciones
Redes Neuronales					
Arboles de decisión ID3, C4.5, C5.0					
Arboles de decisión CART					
Otros árboles de decisión					
Redes de Kohonen					
Regresión Lineal					
Regresión Logística					
K-Means					
A priori					
Naive Bayes					
Vecinos más próximos					
Algoritmos genéticos y evolutivos					
Máquinas de vectores de soporte					
Análisis discriminante multivalente					

Tabla 1: Relación entre técnicas y métodos de minería de datos.

3.2. MINERÍA DE DATOS PARA LA DETECCIÓN DE FRAUDE Y CORRUPCIÓN

Como consecuencia de la realidad actual que enfrentan los sistemas y procesos deben enfrentarse a crecientes volúmenes de información, la minería de datos posee por lo tanto una ventaja teórica sobre las técnicas manuales en la búsqueda de evidencias al evitar la subjetividad de los análisis y optimizar sustancialmente los tiempos requeridos para realizar (Kuna, 2014).

Un claro antecedente en el uso de herramientas computacionales para la detección de fraude (Abbot et al., 1998), donde se testearon un conjunto de modelos para la detección de falsos positivos en transacciones fraudulentas estableciendo la superioridad de árboles de decisión y redes neuronales para acometer esa tarea.

En otro caso de aplicación (Spatis, 2002) se desarrollaron dos modelos orientados hacia la detección de una gerencia fraudulenta en empresas privadas, a través del empleo de regresiones lógicas, representa un caso aplicado en este sentido. Otros estudios que utilizaron minería de datos en la búsqueda de formas de corrupción en la actividad privada han sido relacionadas con el fraude bancario, más específicamente a los vinculados con tarjetas de crédito y aseguradoras de automotores (Foster y Stine, 2004).

A diferencia del caso de la corrupción pública, los casos de corrupción privada compleja, representada por las actividades fraudulenta persistentes en la alta gerencia corporativa, si supone un paradigma en la utilización de técnicas y herramientas de EI. Estos incluyen toda falsificación intencional de estados contables y financieros con la intención de obtener beneficios ilegales (Wang et al., 2006; Wang, 2010), lo cual se hace no solo para evadir o eludir impuestos sino también con el objetivo de confundir a inversores y acreedores.

En este sentido los delitos de corrupción corporativa vinculados generalmente con la alta gerencia, no solo han venido han aumentado considerablemente (Koskivaara, 2004) sino que se han profundizado a partir de la crisis financiera de 2008 y por consiguiente

los requerimientos para detectar, definir, y reportar fraude financiero y contable han aumentado (Yue X. et al., 2007). Antes del agravamiento de la corrupción corporativa global varios autores efectuaron contribuciones que contribuyeran a la lucha de esta problemática compleja de corrupción privada.

Otros autores (Green y Choi, 1997) realizaron una clasificación de fraudes en la alta dirección corporativa a partir del uso de redes neuronales; así como se analizó el trabajo de auditores en la utilización de CAATs a través de sistemas de expertos para una mejor discriminación de riesgo de fraude corporativo en distintos niveles de gestión (Eining et al., 1997).

Respecto de corrupción corporativa interna también se llegaron a construir modelos de detección de fraude interno (Fannin y Cogger, 1998) por medio de la utilización de redes neuronales junto a ratios financieros y variables cualitativas aseverando que su modelo era más eficaz que métodos estadísticos tradicionales.

Otros autores como Kirkos y otros (Kirkos et al., 2007) fueron incluso más allá elaborando una comparación de técnicas de minería de datos en la auditoría corporativa de sistemas; en el cual a partir de los estados contables y financieros de 76 firmas griegas se evaluaron la performance de tres técnicas particulares de EI: árboles de decisión, redes neuronales y redes bayesianas en dos fases, una de entrenamiento y otra de validación, encontrando que las RN eran más eficaces a la hora de determinar las firmas que sufrían fraude con el modelo sin entrenar mientras que las RB mostraron un mejor desempeño con los modelos entrenados en etapa de validación.

Clasificación de técnicas de minería para la detección de fraude

Diversos autores han especificado una clasificación de la corrupción privada la cual se manifiesta en la forma de diversos fraudes financieros, en (Ngai et al., 2011) puede apreciarse una clasificación exhaustiva de estos últimos los cuales pueden comprender fraudes bancarios, tarjetas de crédito, blanqueo o lavado de dinero y fraude hipotecario. Los fraudes llevados a cabo por aseguradores comprenden un amplio universo y

comprenden las estafas por cobro fraudulento de seguros por cosechas, seguros de salud, automóviles etc. Finalmente, quizás la categoría más compleja de corrupción privada sea la corporativa, fraudes financieros más complejos que involucran la falsificación de información corporativa, la especulación financiera, etc.



Figura 4: Clasificación de técnicas de explotación de la información para la detección de fraude financiero (Ngai et al., 2011).

La Figura 4 presenta una clasificación de fraudes y técnicas de minería para combatirlo. Similarmente otros autores (Sowjanya y Jyotsna, 2013), hicieron también lo propio en pos de confeccionar una taxonomía abarcativa de la aplicación de minería de datos en la detección de fraude, como se presenta en la Figura 5.

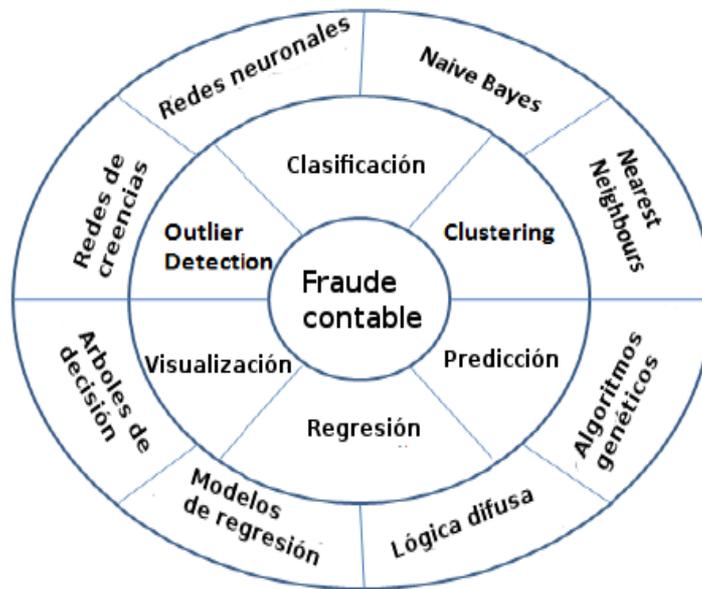


Figura 5: Algoritmos y técnicas de minería de datos para el descubrimiento de fraude fiscal (Sowjanya y Jyotsna, 2013).

Ahora considerando propiamente al fenómeno corruptivo a nivel público y sistémico, un esfuerzo reciente se despliega en el trabajo de Ransom (2013) en uno de los pocos trabajos en el campo abordando el fenómeno de la corrupción como caso de estudio, midieron los niveles de discusión sobre corrupción usando técnicas de minería de texto (*Text Mining*) y extracción de la información (*Information Extraction*) de un cuerpo de datos extraído de noticias y reportes sobre corrupción en 215 países en un lapso de cinco años, lo que les permitió discutir que tan prominente era el debate público sobre corrupción en cada nación. En el estudio los autores afirman de la utilidad de la minería de datos para la diseminación de ideas y conceptos en pequeñas organizaciones civiles que luchan contra la corrupción, reivindicando la utilidad de la minería de datos como herramienta en ciencias sociales, el monitoreo y la investigación científica.

Por otra parte, otros autores (Huysmans et al., 2006; Huysmans, et al. 2008) aplican técnicas de minería en un estudio transversal en cross entre países buscando relacionando variables macroeconómicas a niveles percibidos de corrupción en dos pasos: usando mapas auto-organizados de Kohonen (SOM) *a priori* y maquinas de

vector soporte *-Support Vector Machines-* (SVM) supervisadas *a posteriori* son entrenadas para pronosticar niveles de corrupción futura para cada país, y volcando *a fortiori* tales pronósticos nuevamente en mapas auto-organizados para comparar distintos modelos de comportamiento.

En los casos anteriores, aunque ambos representan esfuerzos rigurosos en cuanto a la búsqueda de la verdad, los mismos representan un enfoque informativo para la sociedad, pero de alcance netamente unidimensional y acotado, sin considerar la actividad del Ciudadano cívico y sus organizaciones civiles que buscan una herramienta para “auditar” a sus conciudadanos representantes.

Sin desmedro de las restantes técnicas de minería de datos, las técnicas de detección de outliers o valores anómalos ha sido escasamente implementada para la solución escenarios de fraude, y por supuesto mucho menos aún, dentro del estudio de la corrupción en el ámbito público. Un caso de aplicación en este sentido para el descubrimiento de corrupción privada en el uso específico de técnicas de detección de datos anómalos fue materializado sobre transacciones de tarjeta de crédito (Aleskerov et al., 1997) de manera de buscar algún tipo de patrón corrupto relacionado con la existencia de fraude.



Figura 6: Técnicas de EI usados para la detección de corrupción pública (Elaboración propia).

En la Figura 6, podemos plantearnos entonces la cuestión de si es posible la aplicación de técnicas de detección de outliers ahora para la detección de fraude en el ámbito

público, aspectos los cuales poseen un alcance de trascendencia social e institucional particular con efectos profundos sobre la sociedad representada.

Los campos anómalos u outliers -siguiendo a Hawkins (1980)- se definen como un dato que por ser muy diferente a los demás pertenecientes a un mismo conjunto de datos, i.e.: una base de datos contenedora de tales campos, puede considerarse que fue creada por un mecanismo diferente; lo que, en el descubrimiento de tales mecanismos, radica el conocimiento latente en cada base analizada.

3.3. MÉTODOS PARA LA DETECCIÓN DE OUTLIERS

La búsqueda de valores anómalos en datos se remonta a más de dos siglos atrás donde ante la presencia de datos discordantes se procedía a excluirlos del resto de la muestra (Boscovich, 1757). A lo largo del tiempo fueron evolucionando en técnicas más complejas y eminentemente luego del desarrollo de técnicas de cómputo y la diseminación del computador.

Como estudiaron varios autores (Chandola et al., 2009), las técnicas de detección de datos anómalos poseen componentes principales asociados con cualquier técnica o método de detección de outliers como se describe a continuación en la siguiente Figura 7.

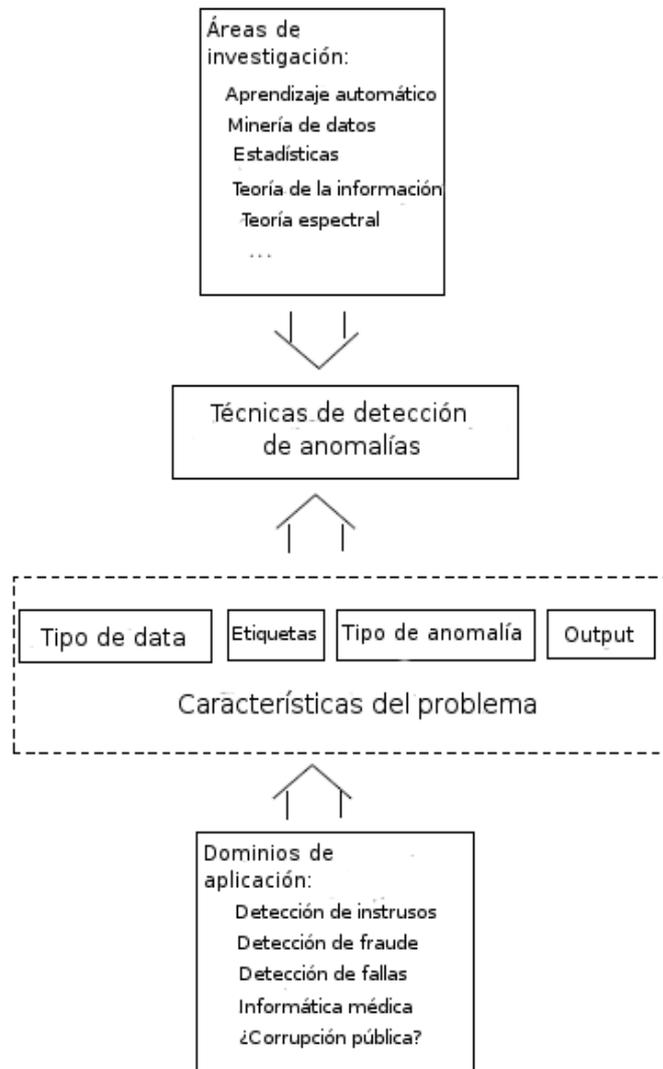


Figura 7: Componentes de las técnicas de detección de anomalías (Elaboración propia en base a Chandola et al., 2009).

Como se desprende de la Figura 7, la minería de datos viene a ser simplemente un área más potencial más para el descubrimiento de conocimiento científico, puesto que al mismo tiempo considera las estadísticas, aprendizaje automático, teoría de la información y teoría espectral como áreas de investigación útiles para su uso.

Estas técnicas a su vez, para abordar la detección de outliers, presentan tres enfoques básicos para su resolución (Hodge & Austin, 2004; Chandola et al., 2009); como se expone a continuación del siguiente cuadro extraído de Kuna (2014).

Enfoque tipo 1:	Detección de outliers con aprendizaje no supervisado.
Enfoque tipo 2:	Detección de outliers con aprendizaje supervisado.
Enfoque tipo 3:	Detección de outliers con aprendizaje semi-supervisado.

Tabla 2: Enfoques para el problema de la detección de outliers (Kuna, 2014).

En la Tabla 2, se describen los tres enfoques posibles que pueden adoptar técnicas de detección de datos anómalos y ruido. En consideración a la carencia de procesos y algoritmos dirigidos a la detección de datos anómalos en resolución de la problemática de ética pública en las sociedades, como se expone en la Figura 6, se pone de relieve una ventana de oportunidad para la propuesta de búsqueda y experimentación tanto novedosa como oportuna de procesos desarrollados recientemente para la la detección de campos anómalos y ruido, útiles para la detección de datos anómalos y ruido en BBDD civiles y públicas; estas, harto sensibles al descubrimiento de nueva información y conocimiento de posibles procesos corruptivos que afectan el bienestar y el tejido social. En este sentido, y a sabiendas que las DDJJ se conforman habitualmente por datos alfanuméricos, se descubre con gran potencialidad de aplicación hipotética, la utilización de los procedimientos híbridos III y IV de detección de datos anómalos desarrollado recientemente por Kuna (2014), los cuales se identifican como idóneos por las siguientes razones:

- Son procedimientos desarrollados recientemente y representan el estado del arte en lo que hace a la detección de campos anómalos y ruido.
- Son procedimientos óptimos para la detección de ruido en BBDD alfanuméricas como generalmente se encuentran caracterizados los datos públicos.
- Son de fácil aplicabilidad y ejecución con los programas de minería de datos disponibles actualmente.

La idoneidad de estos procedimientos son propuestas como solución hipotética para un caso de contralor para la mejora de los datos públicos y cívicos de una población. Como se procede a continuación donde se despliega una posible resolución y aplicación hipotética de contralor civil.

4. SOLUCIÓN PARA UN CASO DE CONTRALOR

En este Capítulo se hecha luz sobre los remedios disponibles para la lucha anticorruptiva. En la primera sección se abre el camino a la explicación de los mecanismos y sistemas político sociales que promueven el desglose de escenarios corruptivos complejos para luego ahondar en técnicas y procesos de minería de datos que soporten los primeros. La exploración de los sistemas de DDJJ como herramientas para la prevención del fenómeno para finalmente proponer una aproximación metodológica para para detección de campos anómalos y ruido en DDJJ en la última Sección, i. e. la posibilidad de proponer el uso de metodologías y procesos de minería de datos para arribar a mejores procedimiento de contralor en administración pública.

Hasta ahora, el abordaje de los trabajos científicos sobre corrupción que emplearon minería de datos lo hicieron propugnando una unidireccionalidad del conocimiento sin considerar la realidad digital contemporánea en la que se encuentra situada la ciudadanía actualmente. Sin considerar tal situacionalidad y potencial digital de la ciudadanía, como tantos otros trabajos (Ransom, 2013; Huysman et al., 2008, 2006); desde las bases usadas para realizar las técnicas de minería de textos en uno, como los índices de percepción de corrupción en el otro, puede llegar a ser discutidas aspectos de acuerdo a las fuentes utilizadas así como su inherente subjetividad, los análisis intra-países también pueden resultar dificultosos de interpretar puesto que al ser la corrupción un fenómeno complejo, este puede adquirir distintos aspectos culturales endógenos para cada país.

Este trabajo pretende abordar el problema de la corrupción desde una perspectiva sistémica diferente, de manera de no solo ofrecer una herramienta para el mayor control a la ciudadanía y organizaciones sociales, sino también poner de relieve la situación patrimonial del alto funcionario publico y así contribuir a poner en valor el ejemplo ético y político como herramienta para la prevención del comportamiento corrupto.

4.1. MEDIDAS PARA LA PREVENCIÓN DE LA CORRUPCIÓN

Prevenir la corrupción y promover la transparencia en la gestión pública, implica desarrollar una mirada compleja que integre políticas y prácticas punitivas con políticas y prácticas preventivas (Gómez & Bello, 2009), un paradigma reciente que se aleja del enfoque tradicional que se apoyaba fundamentalmente en la penalidad de agente corrupto.

La respuesta penal punitivo al problema de la corrupción como única solución demostró ser limitada y poco sustentable en el largo plazo (Canavese 2005, Gómez & Bello 2009) por lo que es necesario desarrollar una mirada compleja que combine políticas y prácticas punitivas con prácticas preventivas de los hechos corruptivos.

La propuesta economicista era presentada generalmente en términos probabilísticos y de utilidad. En este enfoque pueden destacarse la necesidad de reducción de los incentivos a que el funcionario actúe deshonestamente. Siguiendo a Becker (1968) este sugiere aumentando la probabilidad de que el funcionario sea descubierto. Por otra parte el mismo autor sugiere la necesidad de incrementar la satisfacción moral de comportarse honestamente y acrecentar el salario de los funcionarios. Otros autores de este enfoque (Canavese, 2005) ya advirtieron de la inutilidad de la penalidad como solución a este problema social complejo. La principal deficiencia de este enfoque es que se encuentra restringido a una visión asignativa de agentes económicos maximizadores de utilidad rentística sin un capital social haciendo limitada su aplicación real.

Pero la lucha contra la corrupción es también una obligación para los Estados nacionales, la Convención de las Naciones Unidas Contra la Corrupción (Naciones Unidas, 2004; Raigorodsky y Geler, 2007), de la cual Argentina suscribió el acta que afirma que cada Estado procurará [...] *establecer y fomentar prácticas eficaces encaminadas a prevenir la corrupción*. Descartado los enfoques anteriores, las políticas anti-corrupción deben recaer en la prevención.

Según el manual de medidas prácticas contra la corrupción de las Naciones Unidas (Naciones Unidas, 1993; Gómez y Bello, 2009) se consideran centrales los tres siguientes aspectos para acometer dicha tarea:

- El fortalecimiento y la construcción institucional.
- La prevención.
- Toma de conciencia de la ciudadanía.

La primera referente al **fortalecimiento y construcción institucional**, comprende la creación de agencias anti-corrupción; un ombudsman o defensor del pueblo; el fortalecimiento del Poder Judicial, la rendición de cuentas precisa y oportuna; códigos de conducta; comités de integridad nacional o comisiones anti-corrupción; encuentros de integridad nacional para desarrollar planes de acción; fortalecimiento de gobiernos locales.

En segundo lugar, **la prevención**, hace a las declaraciones de activos y pasivos de funcionarios públicos como un instrumento para ello así como la creación de una autoridad de monitoreo internacional para la transparencia de los contratos del sector público en las transacciones comerciales internacionales; islas de integridad y pactos de integridad; grupos de coordinación de información: cooperación anti-corrupción del sector privado, reducción de la complejidad en los procedimientos y de la discrecionalidad.

Finalmente **la toma de conciencia de la ciudadanía**, la que se logra a través del acceso a la información pública; movilización de la sociedad civil a través de la educación pública; planes de acción anti-corrupción; capacitación de la prensa y periodismo de investigación.

En este sentido especialistas en el campo ya habían manifestado también de la importancia de los dos últimos puntos dado que hacer operativo el derecho a la información pública en términos de prevención de la corrupción es proporcionar a los Ciudadanos las herramientas para ejercer sus derechos y controlar la acción del Estado

(Raigorodsky y Geler, 2007; Naciones Unidas, 2004). Visión integral que hasta puede entenderse como análogamente compatible a la propuesta a la de OGP, el cual estipula los siguientes cuatro aspectos clave que aseguran un gobierno abierto (OGP, 2012), como se desprende del siguiente cuadro.

Condiciones para un gobierno abierto y participativo			
Transparencia fiscal	Acceso a la información	Divulgación patrimonial de funcionarios	Compromiso Ciudadano

Tabla 3: Aspectos de los gobiernos abiertos (OGP, 2012).

En la Tabla 3, se presentan los criterios básicos de gobierno abierto los cuales son necesarios para incrementar la responsabilidad de la administración pública, fortalecer el compromiso Ciudadano y luchar contra la corrupción (OGP, 2012).

4.1.2 LOS SISTEMAS DE DDJJ PATRIMONIALES COMO MECANISMO DE PREVENCIÓN DE LA CORRUPCIÓN

La revelación y divulgación del patrimonio de los funcionarios oficiales es uno de los cuatro pilares para la elegibilidad de un gobierno abierto (OGP, 2012). Estos sistemas de información se encuentran representados en nuestro país por los sistemas o regímenes de declaraciones juradas; los cuales, poseen tres funciones básicas para lo cual fueron implementados (Gómez & Bello, 2009):

- Controlar la evolución patrimonial de los funcionarios de la función pública para prevenir enriquecimiento ilícito y otros delitos de corrupción.
- Detectar y prevenir conflictos de intereses e incompatibilidades de la función pública.
- Como mecanismo de transparencia y prevención de la corrupción pública.

Por lo tanto, si a esas políticas preventivas en los cuales se incluyen los regímenes y sistemas de declaraciones juradas patrimoniales como herramienta fundamental; se

combinan con la utilización de procedimientos y técnicas de minería de datos, se podrá fortalecer aún más el poder preventivo de la herramienta institucional como factor de la herramienta computacional.

En el sentido preventivo todo régimen de DDJJ de funcionarios públicos representan una herramienta que posibilita:

- El control del adecuado cumplimiento de las funciones públicas que desempeñan los funcionarios.
- Prevenir el desvío de sus deberes éticos.
- Corregir incumplimientos detectados.

Los mismos son documentos donde se expone, la variación patrimonial de oficiales públicos durante el desempeño de sus funciones. Sus antecedentes laborales -en especial aquellas relaciones contractuales mantenidas en forma simultanea en el desempeño del cargo público-; así como relaciones laborales que hayan cesado en un tiempo relativamente breve anterior a la toma de funciones.



Figura 8: Sistema de DDJJ patrimoniales (Gómez y Bello, 2009).

Como se expone en la Figura 7, los sistemas de DDJJ se componen de un universo de funcionarios públicos con sus atributos correspondientes, las DDJJ *per se*, y la cantidad y calidad de la información como output. Estos sistemas constituyen una de las herramientas centrales para la prevención y el combate de la corrupción, ya que representan instrumentos que permiten mejorar significativamente los niveles de control

sobre los funcionarios así como las vías para establecer su responsabilidad (Raigorodsky y Geler, 2007). Un sistema de declaraciones juradas, por otra parte, además que aumenta el costo-oportunidad de acometer conductas desviadas así como mecanismo de transparencia y prevención de la corrupción (Ferro G., Giupponi L., Gómez, N. 2007).

Siguiendo uno de autores más relevante sobre el fenómeno de la corrupción (Klitgaard, 1992) el mantenimiento de los estándares de comportamiento ético por parte de los funcionarios públicos dependerá de la interacción de tres factores:

- Honestidad personal
- Nivel de ambición
- Sentimientos de pertenencia e integración al grupo

Un sistema o régimen de declaraciones juradas tiene una incidencia profunda por sobre estos factores personales puesto que, además de aumentar el costo-oportunidad de acometer conductas desviadas, permite la satisfacción moral del funcionario de cumplir con el deber de declarar su patrimonio y así rendir cuentas a la ciudadanía (Ferro, Giapponi y Gómez; *up supra*; Becker, 1968).

Un sistema de DDJJ abierto y digitalizado disponible para la ciudadanía en acceso real, contribuirá a la reducción de la corrupción, en cuanto mayor transparencia (mediante el incremento del acceso a la información pública), reducción de la brecha digital cívica (mediante la (auto)participación ciudadana en el contralor civil), y una red isonómicamente neutral³ al proporcionar al Ciudadano la posibilidad de usar su tiempo conectivo cívico y digitalmente para la construcción de su propio proyecto fenoménico (López-Pablos, 2015a).

En Argentina según la Ley Nacional de Ética en la función pública (Ley Nacional nro. 25.188), están obligados a realizar declaraciones juradas patrimoniales, los funcionarios con categoría de Director o equivalente, los diputados y senadores del Poder

³ De isonomía o isonómico: igualdad ante la ley, proveniente del marco civil de internet brasilero que busca la no discriminación en la regulación de internet (Vercelli, 2014).

Legislativo, los Jueces y Secretarios del Poder Judicial, el Presidente de la Nación, los Ministros, Secretarios y Subsecretarios del Poder Ejecutivo Nacional entre otros; considerándose falta grave el incumplimiento de la presentación de la declaración jurada; *i. e.*, que un funcionario puede recibir una condena penal si falsea u omite datos en sus DDJJ.

4.2. PROPUESTA DE APROXIMACIÓN METODOLÓGICA PARA LA DETECCIÓN DE RUIDO Y ANOMALÍAS EN DD.JJ

En este Capítulo se propone variantes metodológicas híbridas para la detección de campos anómalos y ruido en BBDD de DDJJ. Siendo que las escasas BBDD de DDJJ disponibles actualmente en Argentina son de naturaleza alfanumérica; en esta sección, se exploran y seleccionan procesos de detección de datos anómalos recientes para dar con una solución aplicable a la tarea propuesta.

4.2.1 PROPUESTA DE PROCEDIMIENTOS PARA LA DETECCIÓN DE OUTLIERS Y RUIDO EN BB.DD.

Generalmente almacenados en bases de datos relacionales, los datos algunas veces son considerados anómalos; *i. e.* diferentes al resto de los datos, ya sea por un error o por negligencia o malintención. La existencia de estas anomalías coarta definitivamente la capacidad de la información para generar predicciones y patrones en los datos, menguando la calidad de los mismos y por consiguiente el conocimiento y las decisiones resultantes de aquellos.

Dado que no es posible detectar campos de datos anómalos utilizando un solo algoritmo para todos los posibles entornos de posibles anomalías que puedan presentarse, se plantea la necesidad de abordar la implementación de procedimientos híbridos para la detección integral de los mismos (Kuna, 2014), de manera de posibilitar la detección de campos anómalos teniendo en cuenta los dominios específicos de su implementación, las características de la bases de datos, la tipicidad su dimensionalidad, etc. Por otro parte, además del problema que plantea la detección de outliers se plantea la existencia

de inliers o datos expuestos como atípicos cuando en realidad no presentan un comportamiento de un verdadero outlier.

Siguiendo procedimientos de explotación de la información para la identificación de datos faltantes, ruido e inconsistencias (Kuna, 2014), se definen reglas para la combinación híbrida de algoritmos en un proceso formal para dicha tarea estipulado en 4 procedimientos. El mismo plantea la utilización de combinaciones de algoritmos para cada uno de los procedimientos, sucintamente se contempla cada uno en el siguiente Tabla.

Procedimiento I:	Set normal, LOF; con enfoque no supervisado y semisupervisado.
Procedimiento II:	Set normal, LOF y K-Means; con enfoque no supervisado.
Procedimiento III:	Set alfanumérico con atributo objetivo, algoritmos de inducción C4.5, LOF y TI; con enfoque no supervisado y supervisado.
Procedimiento IV:	Set alfanumérico sin atributo objetivo, LOF, DBSCAN C4.5, PRISM y K-Means; con enfoque no supervisado y supervisado.

Tabla 4: Algoritmos por procedimiento híbrido para la detección de outliers y ruido (Kuna, 2014).

La Tabla 4, describe la utilización de métodos híbridos con enfoques supervisado, no-supervisado y semisupervisado para la detección de outliers y ruido ya se habían constituido como la mejor alternativa al lograrse la mayor ganancia de información, reducción del espacio de búsqueda y optimización de los procesos (Kuna et al., 2009; Kuna et al., 2010a; Pautch et al., 2011). Investigaciones recientes han determinado que es posible afirmar que la combinación de algoritmos de distinta naturaleza y también la combinación de procedimientos permite detectar outliers con un nivel de confianza mayor al 60%, entendiéndose (Kuna, 2014).

En los procedimientos híbridos I y II expuestos se combinan ambos para detectar campos de outliers en bases de datos numéricas (Kuna et al., 2012; Kuna et al., 2013a);

mientras que el procedimiento III detecta campos de outliers en bases de datos alfanuméricas conteniendo un atributo clase (Kuna et al., 2012c; Kuna et al., 2013b; Kuna, 2014), igualmente al procedimiento IV, el cual también detecta campos en bases alfanumérica solo que sin un atributo target (Kuna et al., 2014; Kuna, 2014).

4.2.2 METODOLOGÍA HÍBRIDA PARA LA DETECCIÓN OUTLIERS

Metodológicamente se propone la utilización de una metodología híbrida para la detección de datos anómalos y ruido. Este proveerá de un procedimiento que sirva para garantizar el control y la calidad de los datos e información que se piense auditar en la esfera pública.

La aplicación de la metodología híbrida para la detección de outliers en bases de datos y ruido seleccionado (Kuna, 2014) puede estandarizarse de acuerdo según dependa de la utilidad de cada procedimiento, un entorno de determinado de BBDD, la algoritmia utilizada sus técnicas y enfoques como se despliega a continuación en el siguiente cuadro correspondiente a la Tabla 4.

Procedimiento	Entorno	Algoritmos y técnicas	Enfoques
I	BBDD numéricas on o sin atributo objetivo	LOF; metadato	Tipo 1 y 3
II	BBDD numéricas con o sin atributo objetivo	LOF; K-Means	Tipo 1
III	BBDD alfanuméricas con un atributo objetivo	C4.5, Teoría de la información; LOF	Tipo 1 y 2
IV	BBDD alfanuméricas que no contienen un atributo objetivo	LOF; DBSCAN; C4.5; RB; PRISM; K-Means	Tipo 1 y 2

Tabla 5: Procedimientos según entorno, algoritmos y enfoque (Kuna, 2014).

Como se desprende de la Tabla 5, los procedimientos I y II permiten detectar outliers en bases de datos numéricas basándose en metadatos de registros considerados normales y en el algoritmo LOF que proporciona un valor local de outlier de donde se define la medida de anomalía para una tupla considerada; sin embargo, para las DDJJ de funcionarios públicos poseen un entorno alfanuméricos, lo que intuye a la utilización de los procedimientos III y IV para la búsqueda de datos anómalos.

4.2.3 PROPUESTA DE MATERIALES Y DATOS PARA LA EXPERIMENTACIÓN

Como propuesta de materiales a ser usados tentativamente en una experimentación hipotética, se propone el uso del Software Rapid Miner como herramienta para la aplicación y experimentación de los flujos de minería correspondientes a los procedimientos desarrollados por Kuna (2014). Por otra parte se indago sobre la existencias de BBDD públicos accesibles y disponibles actualmente en internet mediante la interface del Diario La Nación Data, la cual consta con DDJJ de oficiales públicos Argentinos correspondientes a los tres poderes del Estado, la cual podría ser una fuente solida de aplicación empírica. Dada la extensión y complejidad de las DDJJ disponibles se propone una primera experimentación con DDJJ de inmuebles.

5. CONCLUSIONES

En este trabajo integrador propuso la aplicación de metodología de procedimientos híbridos para la detección de campos anómalos desarrollada por Kuna (2014) sobre DDJJ de inmuebles de oficiales políticos argentinos de primera línea de función, siguiendo metodología reciente y pertinente en la detección de outliers; como medida de refuerzo y mejora en la lucha ciudadana contra la corrupción en administración pública.

Entre los métodos propuestos se destacan los procesos de detección de campos anómalos y ruido desarrollados por Kuna (2014), pues los procedimientos allí contemplados son capaces de detectar datos anómalos en BBDD alfanuméricas como los representados en DDJJ patrimoniales.

La importancia vital de los flujos informativos en la sociedad de información tanto en lo contemporáneo, como en lo porvenir inmediato, mediano y lejano, por más proba que fuera no puede quedar en disponibilidad de acceso abierto solo a una única fuente para el contralor Ciudadano. Este trabajo, como propuesta de aplicación, intentó remarcar la importancia de la minería de datos y las técnicas de explotación de la información como herramientas sumamente útiles para la lucha contra la corrupción tanto desde el contralor estatal como el Ciudadano.

Una hipotética validación empírica rigurosa sobre datos y evidencia con BD de DDJJ reales, podría ser puesta como futuras líneas de investigación a este trabajo. Con la metodología propuesta, el caso de contralor cívico planteado exige la disponibilidad de datos públicos y un sistema de GA disponible en tiempo y forma para su acceso libre en sistemas e infraestructura disponible para cualquier Ciudadano cosa que en Argentina, si bien la ley de de acceso a la información pública ha sido aprobada recientemente, a la fecha de la edición de este trabajo tales plataformas de acceso son en los hechos aún inexistentes.

En última instancia, este trabajo intenta otorgar un marco adicional a la formación cívica del Ciudadano, con el objeto claro de aportar a un capital social más denso y fortalecido donde la corrupción no germine, la construcción de un GA de acceso efectivo; y capaz de incorporar en la cultura endógena de todo aglomerado Ciudadano, un mayor desenvolvimiento cívico en la participación digital que impulse y transforme al contralor civil societario en una actividad bien comunitaria más disponible a todo actor social que decida colaborar con su tiempo y energías en ese sentido. Aquella última, en obvia articulación con la formación cívica formal y tradicional, la cual ya no puede considerarse desligada de la realidad social informacional contemporánea.

BIBLIOGRAFÍA

- Abbott, D. W., Matkovsky, I. P., & Elder, J. F. (1998). An evaluation of high-end data mining tools for fraud detection. *Proceedings of IEEE International Conference on Systems, Man, and Cybernetics*, 3, 2836 -2841.
- Aleskerov, E., B. Freisleben, B. Rao (1997). Cardwatch: A neural network based database mining system for credit card fraud detection. *Proceedings of the IEEE/IAFE, Conference on Computational Intelligence for Financial Engineering (CIFEr)*, 220-226.
- Becker, G. (1968) *Crime and Punishment: An Economic Approach*. *Journal of Political Economy*, 76(2), 169-217.
- Bour, E. A. (2014, Marzo) "Corrupción, un punto de vista estratégico", *ANCE*, CABA. Recuperado de <http://www.anceargentina.org/site/trabajos/Enrique%20Bour%20Marzo%202014.pdf> (Valido al 1/3/2016)
- Boscovich, R. J. (1757). De litteraria expeditione per pontificiam ditionem, et synopsis amplioris operis, ac habentur plura ejus ex exemplaria etiam sensorum impressa. *Bononiensi Scientarum et Artum Instituto Atque Academia Commentarii*, 4, 353-396.
- Canavese A. J. (2005). The Effects of Corruption Organization and Punishment on the Allocation of Resources. *Berkeley Program in Law & Economics. Latin American and Caribbean Law and Economics Association (ALACDE) Annual Papers*. Paper 19. Recuperado de <http://repositories.cdlib.org/bple/alacde/19> (válido al 1/3/2016)
- Chandola, V., Banerjee, A., & Kumar, V. (2009). Anomaly detection: A survey. *ACM Computing Surveys (CSUR)*, 41(3), 15.
- Cisneros C. (2011). El libre acceso a la información pública: una necesidad para el desarrollo de la democracia representativa y sus instituciones. Tesis de maestría, FCJC-UNLP. La Plata. Recuperado de <http://hdl.handle.net/10915/32206> (válido al 1/3/2016)
- Colombo H. L., Antonini S. G., Chong Arias C. D., Istvan R. M., & Peternoster F. M. (2013). Estudio de soluciones tecnológicas para el desarrollo de un modelo factible de participación ciudadana. *Proceedings del XV Workshop de Investigadores en Ciencias de la Computación*, 954-958. Recuperado de: <http://hdl.handle.net/10915/27354> (válido al 1/3/2016)
- Cortina A. (1994). *La ética de la sociedad civil*, Madrid: Ed. Anaya.
- DD.JJ. *Abiertas* (2015). LNDData. Actualizado al 13/1/2014. Recuperado de <http://interactivos.lanacion.com.ar/declaraciones-juradas/> (válido al 1/3/2016)

- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., & Uthurusamy, R. (1996). *Advances in knowledge discovery and data mining*. MIT Press.
- Fanning, K. M., & Cogger, K. O. (1998). Neural network detection of management fraud using published financial data. *International Journal of Intelligent Systems in Accounting, Finance & Management*, 7(1), 21-41.
- Ferro G., Giupponi L., & Gómez, N. (2007, Enero). *Declaraciones Juradas de Funcionarios Públicos. Una herramienta para el control y prevención de la corrupción*. Tecnología informática y gestión pública 2° ed. – CABA: Oficina Anticorrupción. Ministerio de Justicia y Derechos Humanos, 88 p. Disponible en: <http://www.anticorrupcion.gov.ar/documentos/libro%20ddjj%202ed.pdf> (válido al 1/3/2016)
- Foster, D. & Stine, R. (2004). Variable Selection in Data Mining: Building a Predictive Model for Bankruptcy. *Journal of American Statistical Association*, 99, 303-313.
- García Martínez, R., Servente, M., & Pasquini, D. (2003). *Sistemas Inteligentes*. CABA, Argentina: Editorial Nueva Librería.
- Gómez N., & Bello M. A. (2009, Mayo). *Ética, transparencia y lucha contra la corrupción en la administración pública*, Manual para el ejercicio de la función pública, 1ra ed., CABA: Oficina Anticorrupción, Ministerio de Justicia y Derechos Humanos de la Nación. Recuperado de <http://www.anticorrupcion.gov.ar/documentos/Libro%20SICEP%20da%20parte.pdf> (válido al 1/3/2016)
- Green, B. P., & Choi, J. H. (1997). Assessing the risk of management fraud through neural network technology. *Auditing-A Journal Of Practice & Theory*, 16(1), 14-28.
- Heidenheimer, A., Johnston, M., & Le Vine, V. (1989). *Political Corruption: A Handbook*, Transaction Publishers, New-Brunswick.
- Hawkins, D. M. (1980). Identification of outliers. *London: Chapman and Hall*, 11.
- Hodge, V. J., & Austin, J. (2004). A survey of outlier detection methodologies. *Artificial Intelligence Review*, 22(2), 85-126.
- Huysmans J., Baesens B., & Vanthienen J. (2008). “A Data Miner's Approach to Country Corruption Analysis”, *Intelligence and Security Informatics, Studies in Computational Intelligence*, 135, 227-247.

- Huysmans J., Martens D., Baesens B., Vanthienen J., & Van Gestel T. (2006). Country Corruption Analysis with Self Organizing Maps and Support Vector Machines, WISI, LNCS 3917, 103-114. Recuperado de <https://lirias.kuleuven.be/bitstream/123456789/101626/1/> (válido al 1/3/2016)
- Kirkos, E., Spathis, C., & Manolopoulos, Y. (2007). Data mining techniques for the detection of fraudulent financial statements. *Expert Systems with Applications*, 32(4), 995-1003.
- Koskivaara, E. (2004). Artificial neural networks in analytical review procedures. *Managerial Auditing Journal*, 19(2), 191-223.
- Klitgaard R. (1992). Controlando la corrupción, Buenos Aires: Ed. Sudamericana.
- Kuna, H., García Martínez, R., & Villatoro, F. (2009). Identificación de Causales de Abandono de Estudios Universitarios. Uso de Procesos de Explotación de Información. *Revista Iberoamericana de Tecnología en Educación y Educación en Tecnología*, 5, 39-44.
- Kuna, H., García-Martínez, R., & Villatoro, F. (2010a). Pattern Discovery in University Students Desertion Based on Data Mining. In *Advances and Applications in Statistical Sciences Journal*, 2(2): 275-286.
- Kuna, H., Caballero, S., Rambo, A., Meinl, E., Steinhilber, A., Pautch, G., García-Martínez, R., Villatoro, F. (2010b). Avances en procedimientos de la explotación de información para la identificación de datos faltantes, con ruido e inconsistentes. *Proceedings XII Workshop de Investigadores en Ciencias de la Computación*, 137-141.
- Kuna, H., Caballero, S., Rambo, A., Meinl, E., Steinhilber, A., Pautch, G., Rodríguez, D., García-Martínez, R., Villatoro, F. (2010c). Identification of Noisy Data in Databases by Means of a Clustering Process. *Ingeniería de Software e Ingeniería del Conocimiento: Tendencias de Investigación e Innovación Tecnológica en Iberoamérica*, 264-273.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., Steinhilber, A., García-Martínez, R., Villatoro, F. (2011). Avances en procedimientos de la explotación de información con algoritmos basados en la densidad para la identificación de outliers en bases de datos. *Proceedings XIII Workshop de Investigadores en Ciencias de la Computación*. Artículo 3745.
- Kuna, H., García-Martínez, R., & Villatoro, F. (2012a). Automatic Outliers Fields Detection in Databases. In *Journal of Modelling and Simulation of Systems*, 3(1), 14-20.

- Kuna, H., Rambo, A., Caballero, S., Pautsch, G., Rey, M., Cuba, C., García-Martínez, R., Villatoro, F. (2012b). Procedimientos para la identificación de datos anómalos en bases de datos. In *Proceedings of CONISOFT*, 184-193.
- Kuna, H., Pautsch, G., Rey, M., Cuba, C., Rambo, A., Caballero, S., García-Martínez, R., Villatoro, F. (2012c). Comparación de la efectividad de procedimientos de la explotación de información para la identificación de outliers en bases de datos. *Proceedings del XIV Workshop de Investigadores en Ciencias de la Computación*, 296-300.
- Kuna, H., Villatoro, F., & García-Martínez, R. (2013a). Development and Comparison of Procedures for Outlier Detection in Databases. *Computers & Security*. (en evaluación).
- Kuna, H., Pautsch, G., Rambo, A., Rey, M., Cortes, J., Rolón, S. (2013b). Procedimiento de explotación de información para la identificación de campos anómalos en base de datos alfanuméricas. *Revista Latinoamericana de Ingeniería de Software*, 1(3): 102-106.
- Kuna, H., García-Martínez, R., & Villatoro, F. (2014). Outlier detection in audit logs for application systems. *Information Systems*.
- Kuna H. (2014, Marzo). *Procedimientos de explotación de la información para la identificación de datos faltantes con ruido e inconsistentes*, Tesis doctoral, Universidad de Málaga.
- Larose, D. T. (2005). *Discovering knowledge in data: an introduction to data mining*. Wiley. Com.
- Ley Nacional “Ética en el ejercicio de la función pública”, Ley Nacional nro. 25.188, Infoleg. Recuperado de <http://www.infoleg.gov.ar/infolegInternet/anexos/60000-64999/60847/norma.htm> (válido al 1/3/2016)
- López-Pablos, R (2015a, Marzo). “Nociones cibernéticas e informáticas para una actualización de la ecuación de Klitgaard”, Documento de Trabajo MEISI, Escuela de Posgrado UTN. Recuperado de <http://hdl.handle.net/10915/44663> (valido al 1/3/2016)
- Naciones Unidas (2004). *Convención de las Naciones Unidas contra la corrupción, contra la droga y el delito* (CCNUC), Oficina de las NU, Viena. Recuperado de http://www.unodc.org/documents/treaties/UNCAC/Publications/Convention/04-56163_S.pdf (válido al 1/3/2016)
- Ngai, E. W. T., Hu, Y., Wong, Y. H., Chen, Y., & Sun, X. (2011). The application of data mining techniques in financial fraud detection: A classification framework and an academic review of literature. *Decision Support Systems*, 50(3), 559-569.

- Open Government Partnership. (2012, Junio [2014, Marzo]). *OGP: Articles of Government*. Recuperado de http://www.opengovpartnership.org/sites/default/files/attachments/OGP%20ArticlesGov%20March%2019%202014_1.pdf (válido al 1/3/2016)
- Quinlan, J. R. (1993). C4.5: programs for machine learning. *Morgan Kaufmann*.
- Raigorodsky, N., & Geler L. (2007, Agosto). Convención de las Naciones Unidas contra la corrupción: nuevos paradigmas para la prevención y combate de la corrupción en el escenario global, 1ra ed., CABA: Oficina Anticorrupción, Ministerio de Justicia y Derechos Humanos de la Nación. Recuperado de <http://www.anticorrupcion.gov.ar/documentos/Libro%20CNUCC%202ed.pdf> (válido al 1/3/2016)
- Ransom J. (2013, Junio). *Replicating Data Mining Techniques for Development: A Case of Study of Corruption*, Lund University, Master Thesis, Master of Science in International Development and Management. Recuperado de <http://lup.lub.lu.se/record/3798253/file/3910587.pdf> (válido al 1/3/2016)
- Sowjanya, S., & Jyotsna G. (2013, Noviembre). Application of Data Mining Techniques for Financial Accounting Fraud Detection Scheme. *International Journal of Advanced Research in Computer Science and Software Engineering*, 3(11), 717-724.
- Spathis, C. T. (2002). Detecting false financial statements using published data: some evidence from Greece. *Managerial Auditing Journal*, 17(4), 179-191.
- Transparency International (2009, Julio). *The anti-corruption Plain Language Guide*, TI E-Book. Recuperado de http://media.transparency.org/fbooks/pubs/pl_guide/ (válido al 1/3/2016)
- Verchelli (2013). La participación política ciudadana en la era digital, Monterrey: CIES, *Virtualis*, enero-julio 4(7), 115-129.
Recuperado de <http://micampus.ccm.itesm.mx/documents/14896/111549100/virtualis07.pdf> (valido al 1/3/2016)
- Yue, X., Wu, Y., Wang, Y. L., & Chu, C. (2007, Septiembre). A review of data mining-based financial fraud detection research, international conference on wireless communications, *Networking and Mobile Computing*, 5519–5522.
- Wang, J., Liao, Y., Tsai, T. & Hung, G. (2006, Octubre). Technology-based financial frauds in Taiwan: issue and approaches, *IEEE Conference on: Systems, Man and Cyberspace*, 1120–1124.

Wang, S. (2010). A Comprehensive Survey of Data Mining-Based Accounting-Fraud Detection Research. *International Conference on Intelligent Computation Technology and Automation*, vol. 1, pp.50-53, 2010.