

PREDICCIÓN DE POSTULANTES QUE COMETERÁN FRAUDE INTERNO EN UNA COMPAÑÍA CON ALGORITMOS DE APRENDIZAJE SUPERVISADO*

Sergio Espinoza-Montalvo
20142424@aloe.ulima.edu.pe
Universidad de Lima. Lima, Perú

Resumen

El fraude interno es un gran problema para las empresas, ocasionando pérdidas monetarias importantes. Diversas investigaciones han propuesto mejoras al proceso de selección de personal utilizando minería de datos. El presente trabajo propone utilizar la información histórica de postulantes a una empresa para predecir si cometerán fraude durante su estadía. Existen modelos con un nivel de precisión alto, pero que tienen un error de clasificación mayor para encontrar los casos de fraude. Después de diversas experimentaciones, se identifican alrededor de 7 características de este universo que aportan más al modelo. Algunas de estas variables coinciden con variables mencionadas en la literatura encontrada sobre trastornos antisociales. El algoritmo con mejores resultados es una red neuronal convolucional con 80 % de precisión. Se concluye que hay valor en la información de postulantes para determinar si cometerán fraude interno durante su estadía en la empresa.

Palabras clave: aprendizaje supervisado, predicción de fraude, trastorno antisocial, fraude interno

Abstract

Prediction of applicants who will commit internal fraud in a company using supervised learning algorithms

Internal fraud is a big problem for companies since it causes significant monetary losses. Several research studies have proposed to improve the personnel selection process using data mining. The present work suggests to use applicants' historical information in order to predict if they will commit fraud during their working period in a company. There are models with high precision level but with a higher error rate to find fraud. After several experimentations, around seven variables which contribute more to the model were found. Some of these variables match those mentioned in studies about antisocial personality disorder. The algorithm with best results was a convolutional neural network with 80% accuracy rate. It is concluded that applicants' information is important to establish if they will commit internal fraud during their working period in a company.

Keywords: supervised learning, fraud prediction, antisocial personality disorder, internal fraud

* **Agradecimientos.** Se agradece el apoyo de la licenciada en psicología Doris García para analizar las pruebas de Wartegg de los postulantes. De igual manera, la asesoría del profesor José Antonio Taquia en técnicas y herramientas para minería de datos. Además de las recomendaciones de los profesores Juan Gutiérrez, Daniel Cárdenas y Rosario Guzmán para la realización del trabajo.

1. Introducción

El fraude interno es uno de los problemas que está presente en todas las organizaciones y ocasiona pérdidas económicas importantes a las mismas (Kroll, 2012), así como baja reputación, fugas de información, robos de equipos, entre otros (Smith, 2005). En el Perú, aproximadamente, el 80 % de los casos de fraude viene del interior de la compañía y el 22 % de los casos, en 2013, ocasionaron más de 100 mil dólares en pérdidas para las grandes empresas (EY, s. f.).

En vista de las exigentes demandas crecientes de calidad, costo y tiempos en las compañías, la minería de datos es una necesidad para los departamentos de recursos humanos, por el manejo de la información sobre los empleados que se puede explotar, incluyendo los *curriculum vitae*, información de recursos humanos, evaluaciones, entre otros (Horesh, 2016). Reconocer el patrón de comportamiento de los trabajadores en el mercado es uno de los recursos, actualmente, más importantes para las empresas (Rashid, 2016).

En el presente artículo se analizan casos de fraude interno en una empresa de suministros de cómputo en el Perú. Se tuvo acceso a la información histórica de los postulantes para predecir cuales cometerían fraude interno con base en una clasificación con algoritmos de aprendizaje supervisado. El documento consta de las siguientes partes: la sección 2 presentará el contexto relacionado al estado del arte; la 3, la metodología de la investigación; la 4, los resultados obtenidos. Finalmente, las 5 y 6 mostrarán las conclusiones y las recomendaciones para investigaciones futuras.

2. Contexto

Diversas investigaciones han tratado el problema de selección de personal en las empresas utilizando minería de datos. Varshney (2014) utilizó la información de recursos humanos para predecir la aptitud de los postulantes para su cargo, usando SVM (*support vector machine*) y regresión lineal. Una investigación del 2016 utilizó múltiples fuentes de datos sobre los empleados de una organización para poder predecir la aptitud de estos para su empleo en tiempo real; la técnica utilizada fue *ordinal regression clustering*, la cual consiste en aplicar regresión para agrupar datos (Horesh, 2016). En el estudio de Jantan (2011), donde se predijo la aptitud de un empleado para su cargo, el C4.5/J48 demostró ser el más efectivo, con 95,14 % de precisión. En la investigación de Rashid (2016) para predecir el comportamiento de empleados, los árboles de decisión tuvieron un 97,61 % de precisión, mientras que la red neuronal convolucional obtuvo mejores resultados, 98,12 % de precisión.

En investigaciones enfocadas a detectar fraude en transacciones, tales como Bhattacharyya (2011), se tiende a usar el algoritmo *random forest*, donde este mismo tuvo el mejor resultado con 86 % de precisión. Existen diversas técnicas para optimizar los modelos de minería de datos. El experimento de Rashid demostró una mejora en promedio de 1,75 %

de precisión, realizando selección de características con técnicas de *information gain*, *gain ratio*, *oneR* y *fuzzy RST* (Rashid, 2016). El estudio de Horesh (2016) implementó la técnica de terminación de matriz mediante optimización convexa para obtener los datos faltantes (alrededor del 80 %) con buenos resultados. Para optimizar los hiperparámetros, en la investigación de Chang se ha encontrado el uso de métodos Taguchi para predecir el abandono laboral (Chang, 2009).

2.1 Trastornos antisociales

Hay una clara relación entre individuos que cometen actos criminales, tales como fraude interno, e individuos con un trastorno antisocial (Aquino, 2003). De acuerdo al DSM-5 (Diagnostic and Statistical Manual of Mental Disorders, por sus siglas en inglés) (Manual diagnóstico y estadístico de los trastornos mentales), una persona con trastorno de personalidad antisocial (TPA) se caracteriza por incumplir las reglas sociales, mentir seguido, no tener consideración por la seguridad ajena, así como tener una conducta irresponsable y un comportamiento impulsivo (APA, 2013). Algunas prognosis para personas con TPA son suicidio, abuso de sustancias, menor probabilidad de estar casado, violencia, crimen, desempleo y estar sin hogar (Le Corff, 2014; Yavuz, 2016).

2.2 Examen de Wartegg

La prueba de Wartegg (ver figura 1) es un examen proyectivo psicológico que se usa para la selección de personal. Este consiste en ocho campos, donde cada uno tiene un trazo impreso, conocido como estímulo. Los postulantes tienen la libertad de elegir por cual campo empezar y qué ilustrar en cada uno. Hay cuadros en donde se puede escribir el orden en el cual se efectuó la prueba; y, finalmente, se hacen cuatro preguntas sobre los dibujos. Hay un rango de dibujos esperados para cada campo, y basándose en esto, junto con otros factores, se realiza la evaluación.

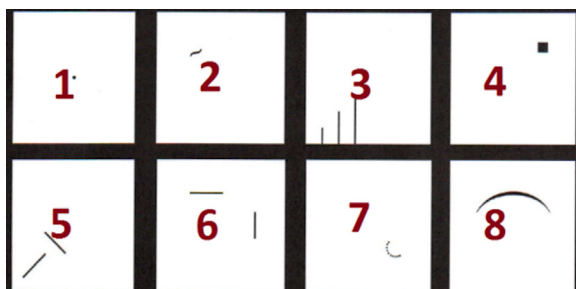


Figura 1. Numeración de los cuadros con los estímulos de la prueba de Wartegg, en ella los postulantes tienen que dibujar en los cuadros de forma libre y responder preguntas.

Elaboración propia

3. Metodología

Para la muestra de la experimentación se utilizaron 63 registros de empleados de una empresa mayorista de suministros de cómputo, los cuales trabajaron entre 2013 y 2016. Dichos registros incluían la hoja de vida del postulante, la hoja de información y pruebas psicológicas realizadas por los mismos. La información fue proporcionada por la empresa de manera física. Adicionalmente, se indicó qué empleados cometieron fraude durante su estadía en la compañía.

Para seleccionar las características de la hoja de vida y de la hoja de información de los postulantes, se usó el conocimiento previo en variables relacionadas con trastornos antisociales. Algunas de estas variables fueron el sexo, los meses promedio entre trabajos, el número de trabajos pasados, el estado civil, entre otras. Para el tratamiento de las pruebas psicológicas, se contó con el apoyo de una psicóloga experta en el examen de Wartegg. La experta dio pautas y entregó material sobre la prueba, el cual se utilizó para crear variables sobre las pruebas psicológicas. Para los 8 campos se colocó una evaluación del 0 al 2, dependiendo de dos criterios de evaluación: uso del estímulo y dibujo del campo. Para los campos 1, 2, 7 y 8 el dibujo debía ser orgánico; para los campos 3, 4, 5 y 6, inorgánico. El uso del estímulo evaluó si el postulante realizaba un dibujo que lo incluyera de forma adecuada. Los criterios para considerar que no se utilizó este de forma adecuada fueron: el dibujo no encapsula o incluye el estímulo y si el dibujo se realiza encima del estímulo. Fallar los criterios mencionados resultaba en una evaluación mejor del campo. Ponderaciones de evaluaciones de los campos se utilizaron para crear variables tales como empatía, objetivos, órdenes, entre otras. Otras variables incorporaban el orden de realización de los campos, complejidad de dibujos y complejidad de descripciones.

Para la experimentación se probó la hipótesis: existen variables relevantes en la información de empleados que permiten predecir fraude interno. Para el procesamiento de la información se utilizaron algoritmos de aprendizaje supervisado encontrados en la literatura, tales como C.45 (Jantan, 2011), redes neuronales (Rashid, 2016) y *random forest* (Bhattacharyya, 2011). Se compararon los resultados obtenidos entre los algoritmos para determinar el más apto para el problema en cuestión. Para optimizar el modelo se emplearon métodos de selección de características (análisis de entropía, análisis de componentes).

4. Resultados

Se realizó el procesamiento de la información con tres algoritmos de aprendizaje supervisado: C.45 (ver figura 2), BRNN (ver tabla 1) y *random forest* (ver tabla 2), los cuales fueron resumidos (ver tabla 3).

En el árbol de decisión se puede apreciar que las dos primeras divisiones se realizaron utilizando las evaluaciones del campo 7 (sensibilidad) y campo 2 (afectividad y emotividad) del examen de Wartegg. De acuerdo a lo indicado por la evaluadora experta, estos campos están relacionados con la empatía. La información del modelo indica que una alta evaluación en estos campos de los postulantes reduce significativamente la probabilidad de que este haya cometido fraude interno. Esto se traduciría a que un postulante con alta empatía tendría poco riesgo de cometer fraude interno, lo que estaría alineado con la información encontrada sobre el TPA.

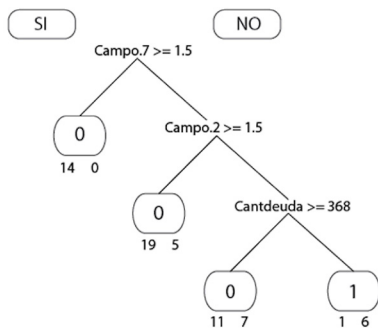


Figura 2. Árbol de decisión obtenido con las características del modelo

Elaboración propia

La red neuronal con regularización bayesiana se realizó con validación cruzada (10 divisiones y 10 repeticiones). El modelo óptimo fue obtenido con 6 neuronas, el cual tuvo un RMSE de 0,4506 y un R cuadrado de 0,1835. El valor de R cuadrado bajo simboliza que el modelo tiene mucha varianza, lo que mostraría que probablemente se necesitan usar características más óptimas (ver tabla 1).

Tabla 1

BRNN obtenida con las características del modelo

Neuronas	RMSE	R2	MAE
1	0,46675	0,1656	0,40528
2	0,45795	0,1822	0,39931
3	0,46050	0,1912	0,40481
4	0,45398	0,1853	0,39732
5	0,45359	0,1860	0,39640

(continúa)

(continuación)

6	0,45061	0,1835	0,39555
7	0,45246	0,1846	0,39647
8	0,45163	0,1856	0,39748
9	0,45213	0,1846	0,39800
10	0,45135	0,1853	0,39702

Elaboración propia

El *random forest* fue generado con 1000 árboles y obtuvo un error OOB (error fuera de caja) de 0,2381, el cual equivale a una precisión de 76,2 %, aproximadamente. Al analizar la matriz de confusión nos podemos percatar que el error de clase, para predecir a las personas que cometieron fraude, es mucho mayor que para predecir a las personas que no cometieron fraude. Esto es preocupante, puesto que simboliza que el modelo está más predispuesto a dar más verdaderos negativos que falsos positivos, lo que no es algo deseado para este tipo de problema.

Tabla 2

Matriz de confusión del *random forest* generado con dos variables

	0	1	Error de clase
0	39	6	13,33 %
1	9	9	50,00 %

Elaboración propia

Comparando los resultados de los tres algoritmos, vemos que, en el contexto del problema, el modelo más efectivo sería el obtenido con *random forest*. Esto se debe a que cuenta con el porcentaje de precisión más alto para encontrar los casos de fraude. El mejor resultado se obtuvo utilizando solo dos variables: trabajos y desempleo (ver tabla 3).

Tabla 3

Resultados del procesamiento de la información con los algoritmos de aprendizaje supervisado

Método	VARIABLES que aportan	Precisión	Precisión (solo para fraudulentos)
RF	Trabajos, desempleo	76,19 %	50,00 %
C.45	Campo.7, Campo.2, Cantdeuda	79,36 %	33,33 %
BRNN	Por determinar	RMSE (0,4647)	Por determinar

Elaboración propia

Para comparar la efectividad de los algoritmos, se realizaron pruebas utilizando diferente número de características (ver tabla 4). Para definir las características a usar, se utilizó el método de análisis de componentes principales (ACP). De este modo, se procesó la información utilizando solo el número indicado de características que aportaban más variación al modelo. La división empleada fue de 52 registros de entrenamiento y 11 de prueba. Se utilizó el mismo valor de semilla para todos los escenarios. Se logró identificar el número de mejores características usadas en la iteración (columna n.º); precisión F o prueba F indica que el porcentaje de precisión es solo para los postulantes que cometieron fraude.

Se puede ver que tanto los algoritmos de C.45 y *random forest* no tuvieron mucha efectividad para poder predecir correctamente la información de prueba, especialmente aquella enfocada en encontrar los casos de fraude. El algoritmo que tuvo los mejores resultados fue la BRNN utilizando 5 registros, donde se obtuvo una precisión de 91 %, aproximadamente, y 50% para los casos de fraude. La selección de características mejoró los resultados hasta cierto punto para cada modelo, donde se puede ver que reduciendo a cierto número de variables los resultados comienzan a empeorar.

Tabla 4
Resultados con diferentes características en los tres algoritmos de aprendizaje supervisado

n.º	C45				Random forest				BRNN			
	Precisión	PrecisiónF	Test	TestF	Precisión	PrecisiónF	Test	TestF	RMSE	Test	TestF	
50	67,50 %	0,00 %	81,82 %	0,00 %	59,62 %	0,00 %	63,64 %	0,00 %	0,5032	72,73 %	50,00 %	
40	66,91 %	0,00 %	81,82 %	0,00 %	59,62 %	0,00 %	63,64 %	0,00 %	0,4882	72,73 %	50,00 %	
30	60,72 %	43,75 %	63,54 %	0,00 %	61,54 %	0,00 %	63,64 %	0,00 %	0,4921	90,91 %	50,00 %	
20	65,03 %	56,25 %	54,55 %	0,00 %	63,46 %	18,75 %	54,55 %	0,00 %	0,4887	90,91 %	50,00 %	
15	62,02 %	50,00 %	45,45 %	0,00 %	69,23 %	18,75 %	54,55 %	0,00 %	0,4919	90,91 %	50,00 %	
10	64,05 %	43,75 %	45,45 %	0,00 %	69,23 %	25,00 %	45,45 %	0,00 %	0,4758	81,82 %	0,00 %	
8	68,25 %	0,00 %	81,82 %	0,00 %	65,38 %	0,00 %	63,64 %	0,00 %	0,4771	81,82 %	0,00 %	
5	66,92 %	0,00 %	81,82 %	0,00 %	51,92 %	0,00 %	63,64 %	0,00 %	0,4689	90,91 %	50,00 %	
3	67,90 %	0,00 %	81,82 %	0,00 %	65,38 %	0,00 %	81,82 %	0,00 %	0,4671	81,82 %	0,00 %	
2	69,63 %	0,00 %	81,82 %	0,00 %	65,38 %	0,00 %	81,82 %	0,00 %	0,4716	81,82 %	0,00 %	
2	64,60 %	0,00 %	81,82 %	0,00 %	59,62 %	0,00 %	63,64 %	0,00 %	0,4636	81,82 %	0,00 %	
2	67,90 %	0,00 %	81,82 %	0,00 %	59,62 %	0,00 %	81,82 %	0,00 %	0,4654	81,82 %	0,00 %	

Elaboración propia

Del análisis (valoración de *random forest*, C.45, PCA y análisis de entropía) se puede determinar que las siguientes fueron las variables más aportantes del modelo.

- Wartegg4 (cuadro elegido en respuesta a la última pregunta del examen de Wartegg, "¿cuál dibujo te pareció más difícil?")
- Desempleo (promedio de tiempo desempleado entre trabajos del postulante. De acuerdo a la teoría de trastornos antisociales [ver sección 2.1], las personas con TPA tienden a estar más tiempo desempleados)
- Trabajos (número de trabajos pasados del postulante. De acuerdo a la teoría de trastornos antisociales [ver sección 2.1], las personas con TPA tienden a rotar más entre trabajos)
- Campo.7 (evaluación del campo 7 del examen de Wartegg utilizando las fórmulas facilitadas por la experta psicóloga, penúltimo campo del examen que está cercanamente relacionado con la empatía)
- Convivientes (número de personas con las que el postulante convive, obtenido de la hoja de información donde se le hacen diversas preguntas al postulante. De acuerdo a la teoría de trastornos antisociales [ver sección 2.1], las personas con TPA tienden más a vivir solos)
- Cantdeuda (cantidad de deuda que tiene el postulante. De acuerdo a la teoría de trastornos antisociales [ver sección 2.1], las personas con TPA tienden a ser más impulsivas y endeudarse más)
- Empatía2 (ponderación de evaluación de los campos de empatía del examen [campos 2 y 8 (unión, armonía y compromiso)])

Todas estas características estarían relacionadas con un trastorno antisocial. Se evaluaron estas variables con un análisis de componentes (ver figura 3). Se puede apreciar que campo.7 y fraude son inversamente proporcionales, lo que concuerda con los modelos hallados y lo esperado; donde a mayor valor de campo.7 (representando una mejor evaluación en el campo para el postulante), menor valor de fraude (equivaliendo fraude = 0 a un postulante que no cometió fraude).

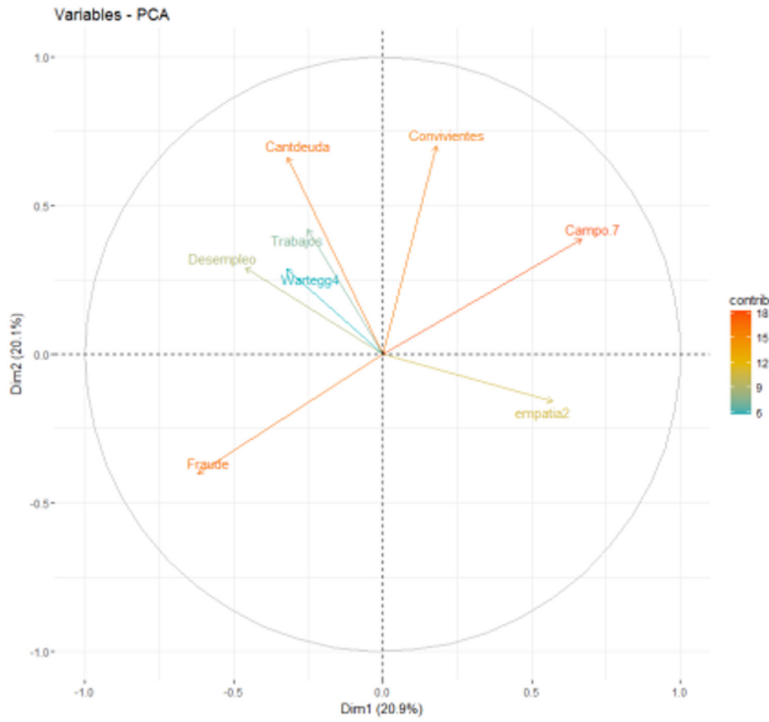


Figura 3. Análisis de componentes principales (ACP) con las variables más relevantes

Elaboración propia

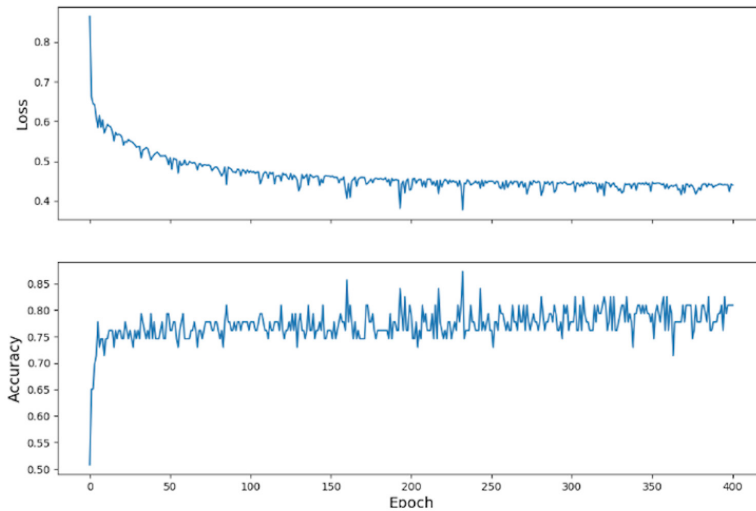


Figura 4. Ratio de aprendizaje de red neuronal entrenada con las mejores características identificadas

Elaboración propia

Finalmente, se entrenó una red neuronal convolucional usando las mejores características identificadas y se obtuvo un porcentaje de precisión de 80 %. El porcentaje de precisión de la red en cada iteración mejoró de forma sistemática (ver figura 4).

6. Conclusiones

Tras el análisis se puede concluir que varias de las características encontradas en los registros de los postulantes están relacionadas con el resultado de que este vaya o no a cometer fraude durante su estadía en la empresa. Sin embargo, la mayoría de modelos tiene más dificultades identificando a los que van a cometer fraude, que a quienes tienen más probabilidad de no hacerlo. Esto explicaría el porqué del error de clase para predecir fraude es tan alto, a pesar de que los valores de precisión global para los modelos estaban entre 70 y 80 %.

Por lo tanto, se concluye que sí se puede predecir si un postulante cometerá fraude interno o no con cierta efectividad. Los resultados parciales muestran que los algoritmos más efectivos para este tipo de modelos serían las redes neuronales. Además, se puede ver que los mejores predictores para este problema están relacionados con trastornos antisociales de acuerdo a lo encontrado en la literatura. Igualmente, con el apoyo de una experta en psicología, se pudieron obtener variables relevantes para el modelo del examen de Wartegg.

7. Trabajos futuros

Se deberían realizar mejoras para reducir el error de clase para detectar fraude, puesto que, como se encontró en la literatura, para este tipo de problemas los falsos positivos son preferibles que los verdaderos negativos.

Uno de los principales problemas es la selección de características. Por un lado, esto se podría solucionar seleccionando, de forma más eficiente, las variables a utilizar en el modelo. Se podrían usar algoritmos genéticos que definan como ajuste del error de clase para predecir el fraude, de modo que se buscaría, específicamente, el mejor modelo que prediga mejor los casos de fraude. Para optimizar los hiperparámetros, se podrían utilizar métodos Taguchi. Por otro lado, no se debería descartar la ayuda de expertos para procesar de manera más efectiva los exámenes psicológicos que se realizan a los postulantes; puesto que se ha evidenciado que de este análisis se pueden obtener características que aportan al modelo. Se recomienda utilizar variables relacionadas a trastornos antisociales según la teoría de TPA, puesto que se puede ver que estas están correlacionadas con la posibilidad de que una persona vaya a cometer fraude o no.

Referencias

- American Psychiatric Association (APA). (2013). *Diagnostic and Statistical Manual of Mental Disorders*. (5th ed.). American Psychiatric Publishing.
- Aquino K., y Douglas, S. (2003). Identity threat and antisocial behavior in organizations: The moderating effects of individual differences, aggressive modeling, and hierarchical status. *Organizational Behavior and Human Decision Processes*, 90(1), pp. 195-208; doi: 10.1016/s0749-5978(02)00517-4
- Bhattacharyya, S., Jha, S., Tharakunnel, K., y Westland, J. C. (2011). Data mining for credit card fraud: A comparative study. *Decision Support Systems*, 50(3), pp. 602-613. doi:10.1016/j.dss.2010.08.008
- Chang, H. Y. (2009). Employee turnover: a novel prediction solution with effective feature selection. *WSEAS International Conference on Computer Engineering and Applications*, 3(6), pp. 417-426.
- EY. (s. f.). Construyendo un ambiente ético. Estudio sobre el riesgo de fraude en el Perú. Recuperado de <http://www.ey.com/pe/es/services/assurance/fraud-investigation---dispute-services/construyendo-un-ambiente-etico-estudio-sobre-el-riesgo-de-fraude-en-el-peru>
- Horesh, R., Varshney, K. R., y Yi, J. (2016). Information retrieval, fusion, completion, and clustering for employee expertise estimation. *IEEE International Conference on Big Data*.
- Jantan, H., Hamdan, A. R., y Othman, A. (2011). Towards applying data mining techniques for talent management. *2009 International Conference on Computer Engineering and Applications, IPCSIT*, 2, pp. 476-481. Singapore: IACSIT.
- Kroll, K. (2012). Keeping the company safe: Preventing and detecting fraud. *Financial Executive*, 28(7), pp. 20-23.
- Le Corff, Y., y Toupin, J. (2014). Overt versus covert conduct disorder symptoms and the prospective prediction of antisocial personality disorders. *Journal of Personality Disorders*, 28(6), pp. 864-872. doi:10.1521/pedi_2012_26_074
- Rashid, T. A., y Asia, L. J. (2016). Improvement on predicting employee behaviour through intelligent techniques. *IET Networks*, 5(5), pp. 136-142. doi:10.1049/iet-net.2015.0106
- Smith, A. D. (2005). *Accountability in EDI systems to prevent employee fraud. Information Systems Management*, 22(2), pp. 30-38.

- Varshney, K. R., Chenthamarakshan, V., Fancher, S. W., Wang, J., Fang, D., y Mojsilović, A. (2014). Predicting employee expertise for talent management in the Enterprise. *KDD 14 Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 1729-1738. New York: ACM. doi:10.1145/2623330.2623337
- Yavuz, K., Şahin, O., Ulusoy, S., İpek, O., y Kurt, E. (2016). Experiential avoidance, empathy, and anger-related attitudes in antisocial personality disorder. *Turkish Journal of Medical Sciences*, 46(6), pp. 1792-1800. doi:10.3906/sag-1601-80