

Uso de minería de datos para la optimización del proceso de reclutamiento en una empresa desarrolladora de software

Abiran López-Ramírez, Raquel Torres-Peralta, Mario Barceló-Valenzuela

Universidad de Sonora, Departamento de Ingeniería Industrial,
Rosales y Blvd. Luis Encinas S/N CP. 83000, Hermosillo, Sonora, México.
abiranlopez@gmail.com, rtorres@industrial.uson.mx, mbarcelo@industrial.uson.mx

Resumen. La minería de datos es una rama de la inteligencia artificial que tiene muchas aplicaciones en la actualidad. Se utiliza principalmente para predicciones fiables de sucesos con base en datos históricos. Este artículo presenta una propuesta para optimizar los procesos de selección de personal en una empresa desarrolladora de software con el fin de acelerar el proceso de evaluación que los ayudará a identificar de manera rápida a los candidatos con mayor potencial.

Palabras clave: Inteligencia Artificial, Minería de Datos, Reclutamiento.

1 Introducción

El capital humano, referente a la inversión en educación y entrenamiento del personal de una empresa que incrementa la productividad de cada individuo [1], es una de las competencias básicas para que las empresas de tecnología mantengan su ventaja competitiva en un área donde rige la adquisición y actualización de los conocimientos. La contratación de personal calificado se ve afectada por el proceso de reclutamiento de las empresas, ya que las evaluaciones necesariamente exhaustivas exigen una secuencia de evaluaciones que hacen que los procedimientos se extiendan por varias semanas.

Los puestos de trabajo requieren personal con más y diversa experiencia por lo que el proceso de selección se hace más complejo. Por lo tanto, los enfoques convencionales de selección de personal que se desarrollan sobre la base de las características del trabajo estático ya no serán suficientes [2]. Con el fin de encontrar a las personas indicadas para un puesto, es vital desarrollar enfoques de selección efectivos. El uso de minería de datos en esta área busca encontrar patrones que distingan a los mejores candidatos de modo que los procesos de selección se ajusten dinámicamente para dirigir los esfuerzos a aquellos profesionales con mayores probabilidades de ser trabajadores exitosos en la organización.

Este documento presenta el marco teórico en la sección 2, los antecedentes y la descripción del problema en la sección 3. La sección 4 describe la propuesta a implementar para la solución del problema y la sección 5 abarca los resultados y beneficios esperados. Finalmente cerramos con una serie de conclusiones.

Abiran López-Ramírez, Raquel Torres-Peralta, Mario Barceló-Valenzuela, *Uso de minería de datos para la optimización del proceso de reclutamiento en una empresa desarrolladora de software*, en: Guillermo Valencia Palomo, José Antonio Hoyo Montaña, Mario Barceló Valenzuela, Alonso Pérez Soltero (Eds.), *Avances de Investigación en Ingeniería en el Estado de Sonora*, pp. 238-243, 2016.

2 Marco Teórico

La selección de personal juega un rol decisivo en la gestión de recursos humanos para determinar la calidad del personal contratado. Investigaciones previas [3-4] han revisado estudios de selección de personal y encontraron que situaciones como cambio en organizaciones, roles y normas de trabajo han influenciado en los procesos de contratación. Mientras que los avances en las tecnologías de información también afectan a la selección de candidatos en la gestión de recursos humanos [5-6]. La aplicación de sistemas de expertos para ayuda en la toma de decisiones en reclutamiento ha crecido [8-9] y las organizaciones están explorando nuevas opciones para mejorar sus procesos de selección de personal y dirigir sus esfuerzos hacia los candidatos óptimos para cada puesto.

Debido a los avances en las tecnologías de información, los investigadores han desarrollado sistemas para la ayuda en la toma de decisiones y sistemas expertos para mejorar los resultados de la gestión de recursos humanos. En particular, la minería de datos es reconocida como una de las disciplinas más sobresalientes, ya que se refiere a la extracción o uso de patrones de una gran base de datos a través de una exploración y análisis automático o semiautomático [10-11]. Estas metodologías se han desarrollado para la exploración y el análisis de grandes cantidades de datos para descubrir patrones y reglas significativas. Los datos, incluyendo información del personal, pueden ser un valioso recurso para el descubrimiento de conocimiento y apoyo a la toma de decisión. En un estudio por la Universidad Nacional Tsing Hua se desarrolló una estrategia para extraer reglas útiles a partir de las relaciones entre los datos de perfil personal y sus comportamientos en el trabajo por medio de árboles de decisión [12]. Los resultados arrojaron reglas a seguir durante el procedimiento de selección para asegurar el seguimiento oportuno a los mejores candidatos. Nuestro enfoque tratará de evaluar distintos algoritmos para identificar los patrones ocultos detrás de los resultados de las entrevistas y exámenes de selección.

3 Problemática

La investigación se desarrollará en una empresa mexicana de outsourcing que se dedica al desarrollo de software y presta sus servicios a compañías internacionales donde los estándares de evaluación son elevados, de modo que la calidad de sus empleados se refleja directamente en los resultados. La elección de candidatos es crucial para alcanzar los objetivos, mantener la relación con el cliente y cuidar el prestigio de la empresa para el crecimiento del negocio. El proceso de reclutamiento es complejo, puesto que consta de 7 etapas por las cuales cada uno de los candidatos debe de pasar: Primer contacto (1), examen de lógica (2), entrevista en inglés (3), entrevista técnico ligera (4), entrevista técnico profunda (5), entrevista cultural (6) y programación conjunta (7). Cada una de las etapas de reclutamiento se califica en base a una escala numérica y cada una de las calificaciones va acompañada con una breve justificación del por qué las personas que estuvieron involucradas en esa parte de la entrevista asignaron ese valor; regularmente es un promedio de 6 personas

diferentes incluyendo parte del equipo al cual el candidato aspira y el reclutador. Cada una de las fases del proceso de reclutamiento requiere una inversión de recursos por ambas partes. El promedio de tiempo invertido a un candidato es 8 horas, esto desde el primer contacto hasta que acepta la oferta. Todo este tiempo se extiende a lo largo de 5 semanas en promedio.

4 Propuesta de solución

Con base a lo anterior y dado que la empresa no cuenta con una forma de saber cuáles de los candidatos pasarán a la etapa final, el departamento de recursos humanos no puede priorizar sus procesos de reclutamiento y eso ha ocasionado que muchas veces pierda a postulantes con mayor experiencia y mejor preparación, pues durante su proceso de aplicación son contactados por empresas de la competencia. Nuestra propuesta consiste en la utilización de un algoritmo de aprendizaje supervisado de inteligencia artificial capaz de detectar los patrones de comportamiento en las evaluaciones de cada uno de los entrevistados en etapas tempranas y predecir cuáles de ellos tiene una mayor probabilidad de ser un desarrollador exitoso dentro de la compañía, para poder así priorizar su proceso de reclutamiento.

La empresa cuenta con un historial de desempeño en evaluaciones durante el proceso de selección para cada candidato. Algunos de los candidatos fueron siendo descartados en diferentes etapas, siendo aquellos que llegaron al final los aptos para ocupar un puesto.

El historial de resultados de los exámenes aplicados a los candidatos se utilizará para entrenar los modelos. Entre los datos que forman parte de la representación de las instancias se encuentran el puntaje de cada evaluación y de forma más granular los aciertos para cada uno de los reactivos de los primeros dos exámenes presentados.

Los candidatos serán clasificados en 5 clases, donde cada clase indica el nivel al que cada uno llegó durante el proceso. Se evaluarán dos modelos de clasificación: el modelo de máquina de soporte vectorial y un clasificador ingenuo de Bayes. Ambos resultados serán comparados con el desempeño de árboles de decisión. A continuación se describen los modelos y el procedimiento de entrenamiento y evaluación.

Los datos personales y características demográficas de los individuos no serán tomados en cuenta durante esta etapa, pues nuestros resultados se basarán exclusivamente en su capacidad para la resolución de problemas.

4.1 Árboles de decisión

Uno de los algoritmos más utilizados para la toma de decisiones a nivel empresarial es el árbol de decisión. Su función es evaluar los grupos de datos y producir reglas que ayuden a la clasificación de una nueva clase (Figura 1a). En nuestra propuesta, los modelos de máquina de soporte vectorial y un clasificador ingenuo de Bayes serán comparados con los resultados de clasificación por medio de árboles de decisión.

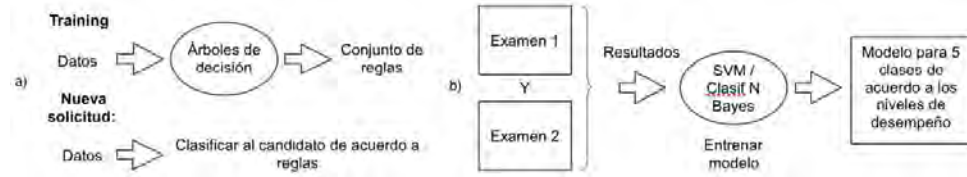


Fig. 1. a) Aplicación tradicional: árboles de decisión. b) Entrenamiento del modelo utilizando una máquina de soporte de vectores (SVM).

4.2 El Modelo de Máquina de Soporte Vectorial

Uno de los modelos propuestos para la identificación temprana de los candidatos óptimos para el puesto consiste en una máquina de soporte vectorial (SVM por sus siglas en inglés). El modelo multiclase (5 clases) será entrenado con el historial de exámenes de los distintos candidatos (Figura 1b) para ser probado con la clasificación de nuevas instancias, esto es, la predicción del nivel al que un candidato puede llegar basándose en los resultados de los dos primeros exámenes.

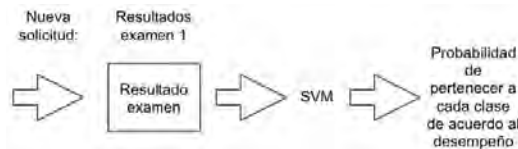


Fig. 2. Clasificación con base a resultados de las evaluaciones en las primeras etapas del proceso de selección de personal.

4.3 Clasificador Ingenuo de Bayes

Al igual de los modelos anteriores, un clasificador ingenuo de Bayes será entrenado con los historiales para evaluar nuevas instancias y predecir la probabilidad de que un candidato llegue a determinado nivel. A pesar de su sencillez, el clasificador ha demostrado su potencia en la clasificación de instancias asumiendo la independencia entre los eventos, que en este caso son las respuestas a las preguntas de los primeros dos exámenes.

Una vez que el candidato presenta sus primeras evaluaciones sus resultados son evaluados por el modelo previamente entrenado. El resultado será un vector con la probabilidad de pertenencia del candidato a la clase de acuerdo a su desempeño (Figura 2). Este resultado permitirá aplicar criterios para la selección del candidato.

5 Resultados y beneficios esperados

El proyecto de investigación evaluará varias opciones para seleccionar un algoritmo adecuado para los datos históricos con los que la empresa de desarrollo cuenta sobre sus candidatos con el fin de ser implementado y que los futuros aspirantes sean sometidos para poder predecir su éxito en el proceso. Nuestra propuesta pretende superar los resultados de un árbol de decisiones, puesto que este necesita instancias que cubran todas las posibilidades en la resolución de un examen, y muchas de las veces los datos no existen. Por otro lado, la implementación de un modelo de máquina de soporte vectorial o un modelo ingenuo de Bayes permitirán la generación de modelos jerárquicos en un futuro para mejorar el desempeño del clasificador. Los beneficios de esta implementación se traducen en una detección temprana de los candidatos más competentes para encaminar esfuerzos a su contratación y evitar el riesgo de perderlo.

6 Conclusiones

El equipo de reclutamiento y la selección de personal juegan un papel que se ve reflejado directamente en la calidad y desempeño de los trabajadores. Actualmente estos procesos son reforzados por técnicas como la ayuda para la toma de decisiones y tecnologías de la información. Un modelo de clasificación para la detección de los mejores candidatos en sus primeras evaluaciones puede revolucionar los procedimientos de reclutamiento haciéndolos más dinámicos y eficientes valiéndose del historial de contrataciones y la información actual de los aspirantes. La minería de datos puede jugar un papel importante para ayudar a los procesos de selección al predecir cuáles candidatos siguen un patrón de éxito y de esta manera identificar a estos elementos claves para darles prioridad sobre otros reduciendo el riesgo de perderlos durante el proceso, mejorando la calidad de los empleados que redundará en mejores productos en la compañía, un aumento en las ganancias y reducción de costos a largo plazo.

Referencias

1. Goldin, C.: Human capital. *Handbook of Cliometrics*, 55-86 (2016).
2. Borman, W. C., Hanson, M. A., Hedge, J. W.: Personnel selection. *Annual review of psychology* 48, 299-337 (1997).
3. Robertson, I. T., Smith, M.: Personnel selection. *Journal of occupational and Organizational psychology* 74, 441-472 (2001).
4. Beckers, A. M., Bsat, M. Z.: A DSS classification model for research in human resource information systems. *Information Systems Management* 19, 41-50 (2002).

5. Kovach, K. A., Cathcart, C. E.: Human resource information systems (HRIS): Providing business with rapid data access, information exchange and strategic advantage. *Public Personnel Management* 28, 275–282 (1999).
6. Liao, S. H.: Knowledge management technologies and applications—literature review from 1995 to 2002. *Expert systems with applications* 25, 155–164 (2003).
7. Hooper, R. S., Galvin, T. P., Kilmer, R. A., Liebowitz, J.: Use of an expert system in a personnel selection process. *Expert systems with Applications* 14, 425–432 (1998).
8. Nussbaum, M., Singer, M., Rosas, R., Castillo, M., Flies, E., Lara, R., Sommers, R.: Decision support system for conflict diagnosis in personnel selection. *Information & Management* 36, 55–62 (1999).
9. Berry, M. J., Linoff, G.: *Data mining techniques: for marketing, sales, and customer support*. John Wiley & Sons, Inc., New York (1997).
10. Chen, M. S., Han, J., Yu, P. S.: Data mining: an overview from a database perspective. *IEEE Transactions on Knowledge and data Engineering* 8, 866–883 (1996).
11. Chien, C. F., Chen, L. F.: Data mining to improve personnel selection and enhance human capital: A case study in high–technology industry. *Expert Systems with applications* 34, 280–290 (2008).
12. Lievens, F., van Dam, K., Anderson, N.: Recent trends and challenges in personnel selection. *Personnel Review* 31, 580–601 (2002).