

¿Qué factores influyen más en el crecimiento de los emprendimientos?

José Cazal¹, María Messina², Esther Hochsztain ²

¹Facultad Politécnica, Universidad Nacional de Asunción.

²Facultad de Ciencias Económicas y de Administración, Universidad de la República.

Resumen

En este artículo se analizan los factores de sostenibilidad de los emprendimientos en el Uruguay. Se identifican los elementos que influyen para que un emprendimiento crezca, así como las motivaciones que tienen los emprendedores para continuar y empezar a generar fuentes de empleo.

Se presenta un caso de estudio basado en una encuesta realizada por Centro de Emprendedurismo CCEEmprende, a los emprendedores participantes programa de apoyo a emprendedores desarrollado por la Facultad de Ciencias Económicas y de Administración de la Universidad de la República del Uruguay.

La metodología aplicada se basa en un proceso de extracción automática de conocimiento empleando minería de datos. Se utiliza la técnica de árboles de decisión como principal modelo de clasificación. Se encara el proceso considerando dos variables explicadas, que se van rotando.

Las variables que participan del estudio son: cantidad de empleados, si el emprendimiento factura, nivel académico de los emprendedores, antigüedad del emprendimiento, cantidad de socios, cantidad de empleados, sigue facturando (si-no), se planifica el trabajo (si-no), cargos definidos en la empresa., aplican procedimientos estandarizados, cuentan con clientes (si-no) y fuente de financiamiento inicial. Las dos primeras son las variables explicadas consideradas.

En primer lugar, se considera como variable explicada la cantidad de empleados con que cuenta el emprendimiento. El modelo obtenido muestra que los emprendimientos que cuentan con mayor número de empleados y más que un año y medio de antigüedad son los que tienen los procesos del emprendimiento definidos y estandarizados. Este resultado indica que ante el crecimiento, las organizaciones necesitan organizarse internamente para una mayor eficiencia y se destaca la importancia de contar procesos estandarizados implementados en la empresa.

En segundo lugar, la variable dependiente pasa a ser si el emprendimiento factura. Los resultados obtenidos muestran que los emprendimientos con más de cinco empleados siguen facturando. Los emprendimientos

que cuentan con menos de cinco empleados, cuyos socios cuentan con estudios universitarios culminados y/o estudios de postgrado ya no continúan facturando. Los emprendimientos con mayor antigüedad siguen facturando independientemente del nivel académico de sus socios.

Los resultados obtenidos brindan información para contribuir a la mejor toma de decisiones en lo relativo a la mejora en los procedimientos de apoyo a emprendedores y emprendimientos. Este estudio se enmarca en un proyecto de largo plazo de investigación de emprendedurismo desarrollado por universidades en Uruguay. Si bien desde la teoría o para otras realidades existe información sobre los factores que colaboran en la determinación del éxito, para la realidad del Uruguay no se identifican proyectos similares.

Palabras claves: emprendedurismo, aprendizaje automático, procesos estandarizados, nivel educativo, antigüedad

1. Introducción

El emprendedurismo hoy en día es de fácil acceso para los jóvenes que inician y deseen obtener una independencia en el ámbito laboral. En Uruguay, organizaciones sin fines de lucro, fundaciones y programas como Emprender y el Programa de Apoyo a Futuros Empresarios (PAFE) dan un aporte interesante a proyectos de emprendedurismo en sus inicios. En el ámbito de la innovación la Agencia Nacional de Investigación e Innovación (ANII) financian proyectos de emprendedurismo evaluados como buenos.

Si bien existe un estímulo en el inicio de los proyectos de emprendedurismo, es importante dar un seguimiento a los proyectos para saber si tienen continuidad en el tiempo, si generan puestos de trabajo y de alguna manera puedan redituarse a la sociedad y a la economía los esfuerzos que se realizan para que los proyectos puedan tener éxito. El emprendedurismo es considerado cada vez más un factor importante para contribuir al desarrollo económico, la generación de puestos de trabajo, la inclusión social, la reducción de la pobreza y el desarrollo de la innovación. Eso ha llevado al interés de políticos, académicos y organismos internacionales en el proceso de creación de nuevos emprendimientos (Messina and Hochsztain, 2015). En este artículo se pretende determinar los factores que permiten clasificar y el rol de los mismos basado en información histórica cuales proyectos de emprendedurismo generan recursos propios y que características tienen los emprendimientos según la cantidad de empleados con que cuenta el emprendimiento.

2. Estado del Arte

En esta sección, se presenta los temas más importantes a desarrollar, primeramente lo referente con emprendedurismo y el método de extracción automática de conocimiento utilizado para extraer información de los datos históricos que es la minería de datos.

2.1. Emprendedurismo

Una organización emprendedora es aquella que asume riesgos, es innovadora y proactiva. Se encuentra en el otro extremo los emprendimientos que son conservadores, que tiene aversión al riesgo, no innova y es reactiva (Covin and Slevin, 1989). Estudio realizado por la Facultad de Ciencias Económicas y de Administración de la Universidad de la República del Uruguay, como precedente mas importante en el Uruguay muestran que los dos elementos más relevantes para anticipar el éxito de un emprendimiento son contar con financiamiento y que, anteriormente, la situación laboral del emprendedor sea trabajador independiente, que se refiere a la situación laboral preexistente del emprendedor (Messina and Hochsztain, 2015). Los emprendedores o socios que se dedican exclusivamente a sus proyectos son los que mayor motivación para el éxito tienen. A nivel de América Latina, se fomenta la economía informal con el emprendedurismo, pero aún se tienen factores como el complicado y largo proceso de formalización de un emprendimiento. En América Latina millones de trabajadores que se desempeñan en este amplio espectro que define a la economía informal, muchos de ellos inmersos en actividades de subsistencia, sin protección social, y donde los derechos laborales brillan por su ausencia (Orsatti and Calle, 2004). Desde el punto de vista de la innovación, Rojas et al. (2013) menciona que el emprendedor debe adquirir una serie de conocimientos que puedan ser aplicados a la vida real y de esta manera innovar en la sociedad a fin de que la empresa donde se ha aprendido ese conocimiento se beneficie de esa innovación y pueda conseguir una ventaja competitiva que sea difícil de imitar por los competidores. Sobre el nivel educativo de los emprendedores, Contín et al. (2007) basado en encuestas realizadas dentro del Proyecto Regional Entrepreneurship Monitor (REM) Navarra 2005, utilizando estadística descriptiva menciona que el 36 % de los emprendedores posee estudios superiores, descendiendo hasta el 22 % en el caso de los empresarios establecidos. Esta evidencia va alineada a los resultados obtenidos en este trabajo, donde los emprendedores establecidos por lo general no tienen estudios terciarios

2.2. Minería de Datos

En la actualidad, literalmente estamos abrumados de datos. La cantidad de datos en nuestras vidas es cada vez mayor y no tienen un horizonte final. Los bajos costos de discos de almacenamiento y los computadores actuales hacen más fácil almacenar algo que tiempo atrás hubiesemos enviado a la papelera. Actualmente nuestras decisiones son registradas electrónicamente todo el tiempo, lo que elegimos en el supermercado, hábitos financieros, nuestros ingresos y egresos. Marcamos nuestro camino a través del mundo, y todo es almacenado en grandes bases de datos (Hall et al., 2011). Con respecto al término de minería se refiere a la extracción de oro de piedras o arena se conoce como la minería de oro en lugar de la minería de roca o arena (Han et al., 2011). La minería de datos se centra básicamente sobre la solución de problemas me-

diante el análisis de los datos ya presentes en bases de datos, encontrar dichos patrones que pueden ser transformados en conocimiento, se define como el proceso de descubrimiento de patrones en los datos de manera automática, o usualmente de manera semi-automática. Esta información puede ser utilizado en el ámbito de los negocios por ejemplo en la ayuda en la toma de decisiones a nivel gerencial, analizar la cartera de clientes, a la competencia, anticipar posibles riesgos en la empresa.

Si bien minería de datos funciona mejor con gran cantidad de datos, (Natek et al., 2013) en su estudio menciona que no tuvo limitaciones con la cantidad pequeña de datos en el uso de herramientas basadas en minería de datos.

2.3. Máquina de aprendizaje.

El aprendizaje automático es un sub-campo de la informática que se desarrolla a partir del estudio de reconocimiento de patrones y la teoría del aprendizaje computacional en la inteligencia artificial. El aprendizaje siempre está basado en un conjunto de datos históricos, observaciones o muestras de una población mayor. En general, la máquina de aprendizaje automático se trata de aprender a hacer mejor en el futuro sobre la base de lo que se vivió en el pasado (Hall et al., 2011), se utilizan los datos para aprender de ellos y también para evaluar el resultado o salida del proceso de aprendizaje, el esquema de aprendizaje tradicional puede ser presentado como en la Figura 1. El énfasis de las máquina de

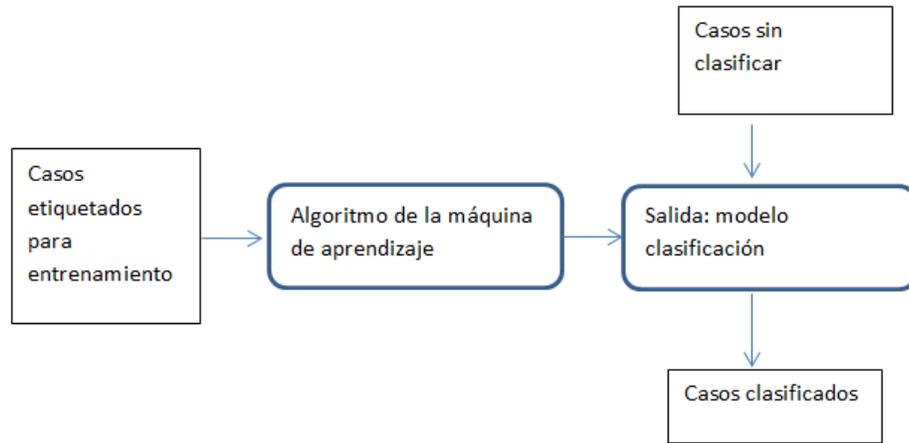


Figura 1: Diagrama de un típico problema de aprendizaje en máquinas de aprendizaje

aprendizaje son los métodos automáticos, en otras palabras, la meta es diseñar algoritmos de aprendizaje que hacen el aprendizaje de forma automática sin intervención humana o asistencia.

Algunos ejemplos de problemas abordados con aprendizaje automático son por ejemplo el reconocimiento óptico de caracteres:(categoriza imágenes escritas a mano a letras representativas), detección de rostros, diagnóstico médico basado en los síntomas, segmentación de clientes, detección de idiomas, predicción del clima, detección de fraudes.

Es importante la interpretabilidad de las reglas de clasificación - predicción, en otras palabras se requiere que los computadores encuentren los patrones o reglas que sean fácilmente entendibles por los expertos humanos.

2.3.1. Técnicas, Algoritmos, Herramientas

Para este trabajo, se utilizó el proceso de extracción automática de conocimiento o por sus siglas en inglés KDD (knowledge discovery in databases). Fayyad et al. (1996) lo define como el proceso que utiliza métodos de minería de datos para extraer lo que es considerado conocimiento de acuerdo con la especificación de medidas y umbrales, utilizando bases de datos junto con cualquier pre-procesamiento requerido, sub-muestreo y transformaciones de los datos. En la Figura 2 expone el proceso consta de 5 pasos:

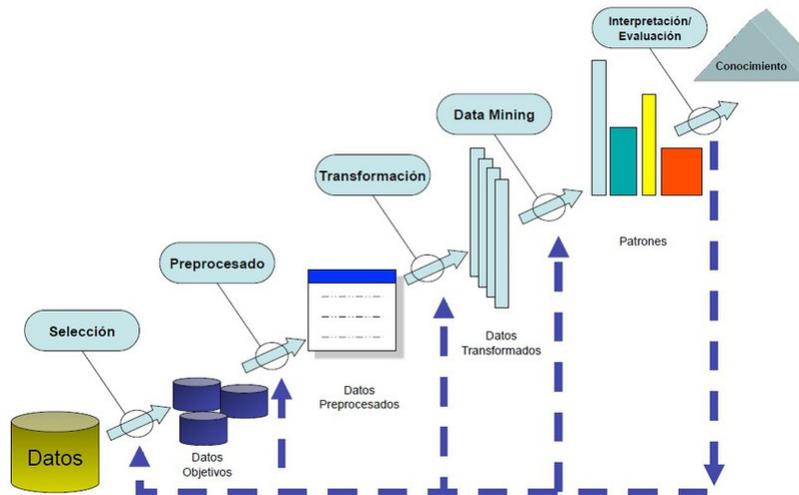


Figura 2: Etapas en el proceso KDD. Obtenido y traducido de Fayyad et al. (1996)

Selección: consiste en seleccionar el conjunto de datos en el cual se efectuará el proceso de descubrimiento.

Pre-procesamiento: en esta etapa se mejora la calidad de los datos, eliminando los valores inconsistentes y tratando los valores nulos o ausentes.

Transformación: consiste en la transformación de los datos reduciendo dimensiones u otras técnicas de transformación.

Minería de datos: en esta etapa se realiza la búsqueda de patrones de interés en una forma particular de representación, dependiendo del objetivo en la minería de datos (clasificación o predicción).

Interpretación/Evaluación: consiste en la interpretación de los patrones encontrados y la evaluación del modelo. Es importante tener un experto en el dominio para poder entender los resultados obtenidos.

En la fase de minería de datos, se utilizan “algoritmos de aprendizaje” para automatizar la búsqueda de patrones en el conjunto de datos analizados. Existen algoritmos cuya salida es bastante intuitiva como las tablas de decisión, árboles de decisión y las reglas. Algunos modelos generados por algoritmos no son interpretables fácilmente como las redes neuronales. En este proyecto se ha utilizado extensamente algoritmos que producen modelos de clasificación cuya salida sean de fácil interpretación por los expertos en el dominio. Los resultados fueron expuestos en formato de árboles de decisión y los resultados fueron de interesante valor para lo que se requiere estudiar en los emprendimientos. El proceso de extracción automática de conocimiento puede realizarse a través de paquetes de software destinados para efectuar análisis de datos, algunos de ellos son SPSS de IBM, gratuitos como R y WEKA. Para nuestro trabajo utilizamos WEKA, que es una herramienta desarrollada por la Universidad de Waikato de Nueva Zelanda. La herramienta es muy intuitiva y nos basamos en el manual de Bouckaert et al. (2015)

3. Estudio de Caso

Se realiza un estudio sobre una muestra de 29 emprendedores encuestados que forman parte del programa Centro de Emprendedurismo CCEmprende, la encuesta fué realizada en el año 2014. En el marco del análisis, el objetivo es identificar que factores influyen en la sostenibilidad del emprendimiento para ello estudia primero las características de los emprendimientos de acuerdo al “número de empleados”. Otro objeto de estudio son los “factores que motivan a que un emprendimiento siga facturando o no”.

3.1. Datos

La encuesta es un formulario de 44 preguntas, y son 29 encuestados, por lo tanto una matriz de 29 filas y 44 columnas.

Los datos son recibidos en bruto, producto de la encuesta. Estos datos pueden presentar inconsistencias, valores mal cargados, etc, factores típicos asociados al relevamiento de datos los cuales deben tratarse. Los datos fueron explorados y

tratados para mejorar su calidad antes de analizarlos propiamente con el método estadístico.

Parte del proceso de análisis es el de identificar los campos que se utilizarán como variables explicativas, de acuerdo al objeto de estudio. A continuación se detalla la metodología utilizada en el proceso de análisis.

3.2. Metodología

La metodología aplicada fue la expuesta en el apartado 2.3.1, donde se detalla paso a paso el proceso de extracción automático de conocimiento.

Primeramente en la fase de *selección*, ya que los datos de la encuesta estaban esparcidos en varios archivos, se procede a unificarlos. Se descartan las columnas o atributos que no son ya de utilidad (como por ejemplo nombre, dirección, teléfono), todos los atributos que no aportan información al objeto de estudio. El conjunto de datos resultante es sometido a una “exploración” de los datos, la exploración consiste en analizar uno a uno los campos de la encuesta, así poder determinar a un nivel mucho más analítico el nivel de relevancia que tienen esos atributos. Además, como toda encuesta se pueden tener datos mal cargado por parte del encuestador, valores ausentes y extremos. Es importante verificar las respuestas para que sean consistentes con el tipo de información que deban contener, por ejemplo al consultar sobre el tiempo que lleva con el emprendimiento, el tipo de respuesta normalmente es un dato numérico que equivale a los meses/años y no encontrar una dirección de correo.

Basado en la relevancia de cada campo y la cantidad de valores nulos, se decide si eliminar el campo o continúa para el análisis posterior.

En la transformación de datos, se redujeron dimensionalidades como por ejemplo, habían campos en la encuesta donde se preguntaban la cantidad de clientes que tenía el emprendimiento a nivel nacional o extranjero. Lo que se realizó fue transformar esos dos atributos a un solo atributo con valores binarios en los que se especificaba únicamente si el emprendimiento posee o no clientes (posee/no posee). Se realizó lo mismo en las fuentes de financiamiento para el inicio del emprendimiento, se redujo a financiamiento propio (ahorros personales) o financiamiento a través de préstamos (a través de préstamos). Atributos numéricos se transformaron a nominales mediante el proceso de discretización, por ejemplo, la antigüedad del emprendimiento en años, se transformó a rangos de valores y así se obtuvieron valores nominales ordinales.

También se identificaron problemas de balanceo de clases en las variables explicadas, por lo tanto se utilizó la técnica SMOTE (Synthetic minority over-sampling technique), lo que hace es crear instancias sintéticas con el objetivo de que la diferencia entre clases no sea muy alta (Chawla et al., 2002).

Al final de la transformación, de los cuarenta y cuatro campos o atributos iniciales, quedaron diez atributos para generar el modelo:

1. Nivel académico.
 - Secundaria completa.

- Universidad incompleta.
 - Universidad completa.
 - Posgrado.
2. Antigüedad del emprendimiento.
- menor a 1,5 años.
 - entre 1,5 y 3 años.
 - mayor a 3 años.
3. Cantidad de socios.
- menor a 3.
 - mayor o igual a 3.
4. Cantidad de empleados.
- igual o menor que 5.
 - mayor que 5.
5. Sigue facturando (Si-No)
6. Se planifica el trabajo (Si-No)
7. Cargos definidos en la empresa.
- No, ninguno.
 - La mayoría no.
 - La mayoría si.
 - Si, todos.
8. Aplican procedimientos estandarizados.
- No, ninguno.
 - La mayoría no.
 - La mayoría si.
 - Si, todos.
9. Cuentan con clientes (Si-No).
10. Fuente de financiamiento inicial (Préstamo-Propio).

Una vez obtenido el conjunto de datos final, se procede a la fase de modelado comenzando al seleccionar la primera variable explicada entre los atributos. El proceso de modelado aplicado en esta metodología se explica en la Figura 3. Para el primer análisis sobre la cantidad de empleados con que cuenta la empresa, se utilizó el atributo número cuatro (Cantidad de empleados) como variable

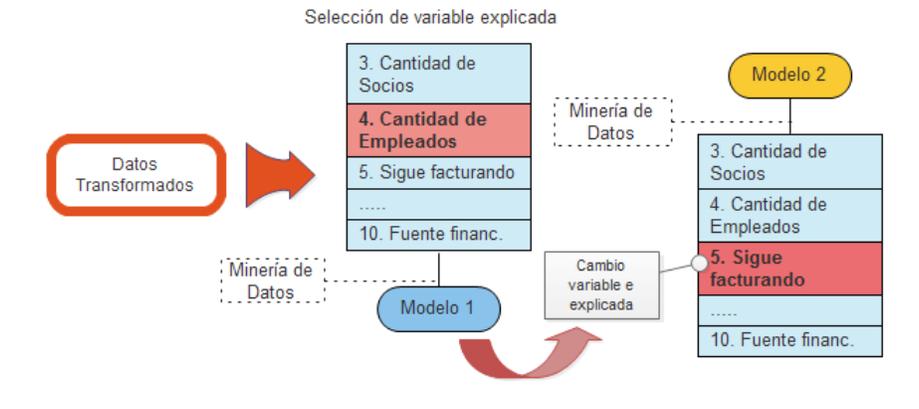


Figura 3: Proceso de modelado, cambiando las variables explicadas

explicada. El resto de las atributos se utilizaron como variables independientes en el análisis.

En la generación del modelo, se utilizó el algoritmo para generar árboles de decisión y luego los resultados fueron analizados por el experto en el dominio. Una vez registrados los resultados obtenidos, se cambió la variable explicada por la número 5 (sigue facturando) y como parte del proceso de análisis, se eliminó el atributo número 8 (Aplican procedimientos analizados) porque el modelo resultante fue inconsistente según los resultados obtenidos, con esto quedaron nueve atributos. El proceso de modelado se ejecutó nuevamente y el resultado arrojó un modelo de clasificación que será expuesto en el siguiente apartado de resultados.

3.3. Resultados

La última fase en el proceso de extracción automático de conocimiento es la evaluación e interpretación de los resultados. La evaluación consiste en obtener el porcentaje de eficacia estimado por cada modelo y la interpretación es realizada por el experto del dominio. El experto del dominio analiza los resultados y lo transforma en conocimiento que se lo utiliza para entender el comportamiento de la variable explicada dada las condiciones de las variables independientes. Los resultados fueron expuestos según las dos variables explicadas consideradas en este estudio las cuales fueron:

- Cantidad de empleados en el emprendimiento.
- El emprendimiento factura o no.

A continuación se detallan los resultados en términos de evaluación e interpretación.

Variable explicada: Cantidad de empleados en el emprendimiento.

El árbol de decisión que se detalla en la Figura 4, representa el modelo obtenido mediante minería de datos, se utiliza como variable explicada la cantidad de empleados en el emprendimiento, la cual posee dos clases: menor o igual a 5 empleados y más que 5 empleados. El modelo tiene una efectividad estimada del 85,7% ¹ Los resultados de la clasificación pueden ser representados en forma

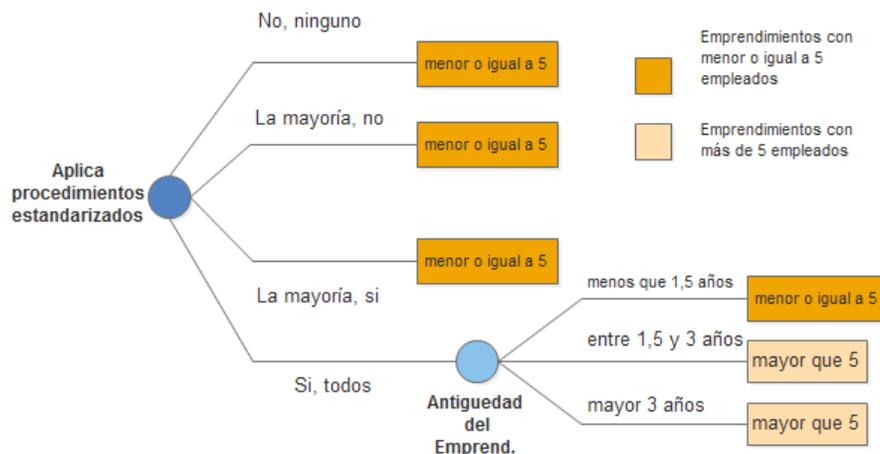


Figura 4: Arbol decisión basado en Cantidad de empleados.

de reglas, las mas destacada es :

- Los emprendimientos tienen mas que cinco empleados y una antigüedad mayor a un año y medio tienen todos sus procedimientos estandarizados.

De esto se pueden derivar una series de conclusiones, primeramente la variable que más discrimina es *Aplica procedimientos estadarizados*, esto destaca la importancia de tener procesos estandarizados.

Variable explicada: El emprendimiento factura o no.

El árbol de decisión que se detalla en la Figura 4, es el modelo obtenido después de cambiar la variable explicada. El modelo tiene una efectividad estimada del 82,5%

Las reglas encontradas más destacadas son:

- Los proyectos de emprendedurismo mayores a tres años de antigüedad si los socios tienen nivel académico de Universidad completa y/o postgrado ya no continúan facturando.

¹El porcentaje es obtenido mediante técnicas de validación cruzada.

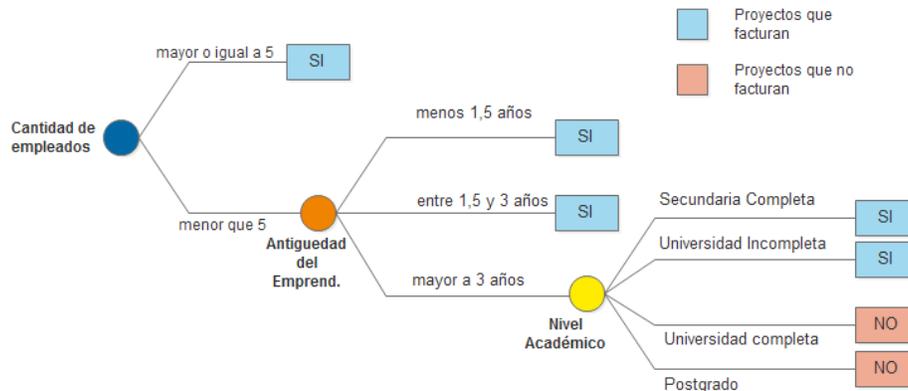


Figura 5: Arbol de decisión si el emprendimiento continúa facturando

- Los proyectos que tengan más de cinco empleados facturan.

4. Conclusiones y Futuros Trabajos

Se cumplió el objetivo de analizar los datos de la encuesta mediante técnicas de minería de datos y se demostró su efectividad. Esto denota que cuanto más empleados y puestos de trabajo generando el emprendimiento, más necesita organizarse ya que los procesos y la planificación aporta un valor agregado en el crecimiento de la organización, destaca la importancia de tener los procesos estandarizados. Incentivar la organización interna de los emprendimientos, no solo cuando este crece, si no desde un principio. Aplicar otras técnicas de minería de datos como reglas de asociación, redes neuronales.

Realizar una encuesta para investigar los motivos por el cual los universitarios dejan el emprendedurismo, la motivación que puede ser económica (porque ganan más como profesionales empleados por otra empresa) o dejan el emprendedurismo por el tiempo que requieren los estudios para poder recibir su título universitario como asistir a clase, estudiar para los exámenes.

Referencias

- Bouckaert, R. R., Frank, E., Hall, M., Kirkby, R., Reutemann, P., Seewald, A., and Scuse, D. (2015). Weka manual for version 3-7-12.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., and Kegelmeyer, W. P. (2002). Smote: synthetic minority over-sampling technique. *Journal of artificial intelligence research*, pages 321–357.

- Contín, I., Larraza, M., and Mas, I. (2007). Características distintivas de los emprendedores y los empresarios establecidos: evidencia a partir de los datos de rem navarra. *Revista de empresa*, 20:10–19.
- Covin, J. G. and Slevin, D. P. (1989). Strategic management of small firms in hostile and benign environments. *Strategic management journal*, 10(1):75–87.
- Fayyad, U. M., Piatetsky-Shapiro, G., Smyth, P., and Uthurusamy, R. (1996). Advances in knowledge discovery and data mining.
- Hall, M., Witten, I., and Frank, E. (2011). Data mining: Practical machine learning tools and techniques. *Kaufmann, Burlington*.
- Han, J., Kamber, M., and Pei, J. (2011). *Data mining: concepts and techniques*. Elsevier.
- Messina, M. and Hochsztain, E. (2015). Factores de éxito de un emprendimiento: Un estudio exploratorio con base en técnicas de data mining (entrepreneurial success factors: An exploratory study based on data mining techniques). *Tec Empresarial*, 9(1):30–40.
- Natek, S., Zwilling, M., et al. (2013). Data mining for small student data set: Knowledge management system for higher education teachers. In *management, knowledge and learning international conference, Zadar*. Retrieved June, volume 1, page 2010.
- Orsatti, A. and Calle, R. (2004). *La situación de los trabajadores de la economía informal en el Cono Sur y el Area Andina: Argentina, Brasil, Chile, Paraguay, Uruguay, Colombia, Perú y Venezuela*, volume 179. Organización Internacional del Trabajo.
- Rojas, R. M., Morales, V. J. G., and Correa, J. A. A. (2013). Análisis de los factores que influyen en el emprendimiento innovador: el aprendizaje organizativo y tecnológico. *Economía industrial*, (388):35–46.