



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

**DESCUBRIMIENTO DE
OPORTUNIDADES DE NEGOCIO
UTILIZANDO REGLAS SIMPLES Y
ÁRBOLES DE DECISIÓN:
APLICACIÓN A START-UP CHILE**

SEBASTIAN IGNACIO ALVAREZ BESOAIN

Tesis para optar al grado de
Magister en Ciencias de la Ingeniería

Profesor Supervisor:
PATRICIO DEL SOL

Santiago de Chile, Julio, 2014

© 2014, Sebastián Álvarez



PONTIFICIA UNIVERSIDAD CATOLICA DE CHILE
ESCUELA DE INGENIERIA

DESCUBRIMIENTO DE OPORTUNIDADES DE NEGOCIO UTILIZANDO REGLAS SIMPLES Y ÁRBOLES DE DECISIÓN: APLICACIÓN A START-UP CHILE

SEBASTIÁN IGNACIO ALVAREZ BESOAIN

Tesis presentada a la Comisión integrada por los profesores:

PATRICIO DEL SOL

MICHAEL LEATHERBEE

JAIME CAICEO

SANTIAGO MINGO

ALFREDO SERPELL

Para completar las exigencias del grado de
Magister en Ciencias de la Ingeniería

Santiago de Chile, Julio, 2014

A mis padres, hermano, familiares y
amigos gracias por su apoyo.

AGRADECIMIENTOS

En primer lugar me gustaría agradecer al profesor Patricio del Sol quien a lo largo de los últimos dos años me ha brindado innumerables oportunidades que me han permitido ir desarrollándome profesionalmente. Además, su apoyo y conocimientos fueron fundamentales para el desarrollo exitoso de esta tesis.

También me gustaría agradecer la participación y compromiso del profesor Michael Leatherbee quien ha enriquecido este trabajo a través de constante *feedback*. Además, Michael presentó un avance de esta investigación en Stanford University siendo posible obtener retroalimentación de primer nivel.

Por último me gustaría agradecer a Rodrigo Pucci con quien trabaje en una primera versión de este trabajo llamada “Reglas Simples para evaluar oportunidades de negocio”, a Miguel Mata y a Carmen Gloria Besoain quienes me apoyaron en la redacción de este trabajo.

INDICE GENERAL

| | Pág. |
|-------------------------------------|------|
| DEDICATORIA..... | ii |
| AGRADECIMIENTOS | iii |
| INDICE DE TABLAS | vi |
| INDICE DE FIGURAS..... | vii |
| RESUMEN..... | viii |
| ABSTRACT | ix |
| 1. Introducción..... | 1 |
| 2. Marco Teórico..... | 2 |
| 3. Start-Up Chile | 4 |
| 4. Datos | 5 |
| 5. Metodología..... | 7 |
| 5.1. Selección de reglas | 7 |
| 5.2. Análisis Estadístico | 12 |
| 5.3. Combinación de reglas | 12 |
| 5.3.1. Construcción del árbol | 14 |
| 6. Resultados..... | 17 |
| 6.1. Análisis Individual | 17 |
| 6.1.1. Factibilidad | 17 |
| 6.1.2. Prototipo..... | 18 |
| 6.1.3. Incertidumbre..... | 19 |
| 6.1.4. Nivel educacional | 20 |
| 6.2. Análisis Estadístico | 21 |
| 6.3. Modelo 1 | 22 |
| 6.4. Modelo 2 | 24 |
| 6.5. Modelo 3 | 26 |

| | |
|---|----|
| 7. Validación de los Modelos | 28 |
| 8. Conclusiones | 29 |
| Bibliografía..... | 31 |
| Anexo A: Cuestionario enviado a los participantes | 35 |
| Anexo B: Output WEKA Modelo 1 | 39 |
| Anexo C: Output WEKA Modelo 2..... | 40 |
| Anexo D: Output WEKA Modelo 3..... | 41 |

INDICE DE TABLAS

| | Pág. |
|--|------|
| Tabla 1: Resultados test Kolgomorov-Smirnov | 6 |
| Tabla 2: Matriz de Costos | 15 |
| Tabla 3: Respuesta de los participantes a la pregunta Q1 | 17 |
| Tabla 4: Aprobación primera regla y selección Demoday | 17 |
| Tabla 5: Respuesta de los participantes a la pregunta "Phase of Development" | 18 |
| Tabla 6: Aprobación segunda regla y selección Demoday | 18 |
| Tabla 7: Respuesta de los participantes a la pregunta Q10 | 19 |
| Tabla 8: Aprobación tercera regla y selección Demoday | 19 |
| Tabla 9: Respuesta de los participantes a la pregunta "What is your Highest Educational Level" | 20 |
| Tabla 10: Aprobación cuarta regla y selección Demoday | 20 |
| Tabla 11: Modelos individuales | 21 |
| Tabla 12: Modelo completo | 22 |
| Tabla 13: Clasificación del árbol de decisión Modelo 1 | 23 |
| Tabla 14: Clasificación de los jueces Modelo 1 | 24 |
| Tabla 15: Clasificación del árbol de decisión Modelo 2 | 25 |
| Tabla 16: Clasificación de los jueces Modelo 2 | 26 |
| Tabla 17: Clasificación del árbol de decisión Modelo 3 | 27 |
| Tabla 18: Clasificación de los jueces Modelo 3 | 27 |

INDICE DE FIGURAS

| | Pág. |
|---|------|
| Figura 1: Clustering del Ranking de Entrada..... | 6 |
| Figura 2: Clasificación mediante un árbol de decisión Modelo 1..... | 23 |
| Figura 3: Clasificación mediante un árbol de decisión Modelo 2..... | 25 |
| Figura 4: Clasificación mediante un árbol de decisión Modelo 3..... | 26 |

RESUMEN

El objetivo de este estudio es desarrollar una metodología que puedan utilizar tanto las empresas de capital de riesgo como los programas gubernamentales orientados al emprendimiento para detectar exitosamente oportunidades de negocio. La metodología se basa en el uso de heurísticas en las etapas de selección preliminar y evaluación, con el objetivo de seleccionar a las mejores oportunidades en lugar de utilizar como criterio de selección la evaluación de expertos.

Para el desarrollo y validación de los modelos se utilizaron datos de Start-Up Chile, un programa de gobierno cuyo objetivo es convertir al país en el polo de la innovación latinoamericana.

En base al análisis de la literatura se escogieron cuatro atributos que estarían relacionados con el descubrimiento de mejores oportunidades, un mejor desempeño y una mayor probabilidad de sobrevivencia a través del tiempo. Estas cuatro reglas se combinaron utilizando un algoritmo de aprendizaje de máquina conocido como árboles de decisión.

Al compararlos en igualdad de condiciones, los tres modelos obtenidos en este estudio tendrían un mayor éxito que la evaluación experta en la detección de potenciales oportunidades.

Palabras Claves: Árboles de decisión, Clasificación de oportunidades, Reglas simples, Heurísticas de selección.

ABSTRACT

The objective of this study is to develop a methodology that could be used by both venture capital companies and government initiatives aimed at entrepreneurship to successfully detect business opportunities. The methodology is based on the use of heuristics at the stages of prescreening and evaluation to select the best opportunities instead of the evaluation of experts as the selection criteria.

For the development and validation of the models we used data of Start-Up Chile, a government initiative aimed to transform the country into a hub of innovation in Latin America.

According to the literature we defined four attributes that would be related to the detection of better opportunities, a better performance and more likelihood of subsistence throughout the time. These four rules were combined using a machine learning algorithm known as decision trees.

If compared on equal terms, the three models developed in this study would be more successful than the detection of potential opportunities done by experts' evaluation.

Keywords: Decision Trees, Opportunities Sorting, Simple Rules, Selection heuristics.

1. INTRODUCCIÓN

La cantidad de propuestas que evalúan las empresas de capital de riesgo y programas de gobierno orientados al emprendimiento y la baja tasa de éxito que tienen ha generado investigaciones en torno a la forma en que se toman las decisiones de inversión, el tiempo que toma este proceso y la información que se considera al momento de evaluar las propuestas. Sin embargo, aún no se ha desarrollado una metodología que se adapte a cada organización y permita explotar la información de experiencias anteriores para: (1) Identificar las heurísticas que utilizan en las etapas de preselección y evaluación; (2) Clasificar oportunidades utilizando combinaciones de las heurísticas detectadas.

El objetivo de este trabajo es elaborar un *framework* que permita clasificar oportunidades, explotando la información de experiencias previas. Para esto se utilizarán técnicas de aprendizaje de máquina, específicamente un algoritmo de clasificación conocido como árboles de decisión.

Los datos utilizados en esta investigación corresponden al programa Start-Up Chile, un programa gubernamental cuyo objetivo es atraer a emprendedores de clase mundial y convertir al país en el polo del emprendimiento latinoamericano.

La base de datos utilizada para la construcción de los modelos comprende 222 observaciones, abarcando las generaciones 4 a 7 del programa. Los participantes se dividen en dos grupos: quienes fueron aceptados para presentar su emprendimiento a inversionistas externos en el Demoday y quienes no fueron seleccionados. Además, se utilizarán los datos de la octava generación para validar el potencial de clasificación de los modelos.

2. MARCO TEÓRICO

La toma de decisiones que realizan los inversionistas de riesgo ha sido ampliamente estudiada en la literatura. Este proceso comprendería entre 5 a 6 etapas secuenciales (Wells, 1974; Tyebjee & Bruno, 1984; Silver, 1985; Hall, 1989) las cuales corresponderían a (1) búsqueda; (2) selección preliminar; (3) evaluación; (4) estructuración del trato; (5) actividades de post-inversión; (6) liquidación, siendo las etapas de mayor interés las de selección preliminar y evaluación.

La etapa de selección preliminar utiliza una serie de reglas no compensatorias (es decir, un valor inaceptable en una regla no puede ser compensado por el valor de otra regla) para seleccionar aquellas oportunidades de negocio cuyo perfil coincide con la estrategia de inversión de la compañía (Grzech, 2009). Las reglas de selección corresponderían al atractivo del mercado, la diferenciación del producto, la capacidad de gestión, amenazas del ambiente y la liquidez de la inversión (Tyebjee & Bruno, 1984; Golis, 1998; Ferris, 2000). Durante esta etapa los inversionistas dan una rápida revisión a las propuestas, decidiendo en menos de 21 minutos si continúan la evaluación o simplemente descartan la propuesta (Fried & Hisrich, 1994).

La etapa de evaluación implica una revisión exhaustiva del plan de negocios y utiliza reglas compensatorias (o sea, una regla cuyo valor no es aceptable puede ser compensado por el cumplimiento de otra regla). Las reglas utilizadas en esta etapa son las mismas que se utilizan en la etapa de selección preliminar pero los aspectos financieros adquieren una mayor relevancia (Riquelme & Rickards, 1992; Hall & Hofer, 1993).

En base a las reglas utilizadas al momento de preseleccionar y evaluar oportunidades de negocio, es posible concluir que se estaría utilizando un modelo similar a las 5 fuerzas de Porter (Van Osnabrugge, 1998; Golis, 1998). El problema de utilizar este modelo como herramienta de evaluación es que fue elaborado para ambientes poco dinámicos, donde es factible posicionarse estratégicamente o utilizar ciertos recursos estratégicos para adquirir una ventaja competitiva a largo plazo (Bingham, Eisenhardt, & Furr, 2011).

Bingham, et. al. (2011) propone el enfoque estratégico de reglas simples para las empresas que se desenvuelven en ambientes dinámicos, donde la ventaja competitiva provendría de detectar exitosa y rápidamente oportunidades de negocios. Las reglas simples corresponden a las directrices que permiten a la compañía posicionarse donde las oportunidades emergen, en otras palabras, sitúan a la compañía al filo del caos.

Este enfoque es el que se utilizara en esta investigación, de manera que al combinar reglas simples o heurísticas de selección obtengamos modelos capaces de clasificar oportunidades de negocio con una mayor eficiencia y objetividad que la que tendría la evaluación de expertos.

3. START-UP CHILE

Start-Up Chile es un programa desarrollado por el Gobierno de Chile, ejecutado por la Corporación de Fomento de la Producción (CORFO) cuyo objetivo es atraer emprendedores de clase mundial para que estos inicien sus negocios en Chile, transformando al país en el polo del emprendimiento latinoamericano.

El programa se desarrolla tres veces al año y recibe aproximadamente 800 postulaciones validas por generación. Estas postulaciones son evaluadas por la consultora YouNoodle, midiendo el potencial de los proyectos en cuatro dimensiones: (1) Productos y Servicios; (2) Mercado; (3) Capital Humano y (4) Capital Social. Los cuatro aspectos son promediados y en base a esta recomendación la CORFO realiza la selección final de los 100 proyectos con mayor potencial.

El programa entrega a los equipos seleccionados Visa por un año, acceso a mentorías durante el transcurso del programa y un monto equivalente a US \$40.000. Una vez finalizado, la CORFO selecciona a los 15 mejores proyectos, quienes presentan sus proyectos a inversionistas extranjeros en el Demoday.

4. DATOS

Los datos corresponden a *startups* seleccionados por la CORFO para participar en el programa y se basan en una muestra de dos partes. La primera parte corresponde a la información que entregaron los participantes al momento de postular al programa, mientras que la segunda corresponde a una encuesta (Anexo A) que se envió a los participantes para evaluar las reglas simples.

Los datos abarcan las últimas cuatro generaciones del programa, por lo que la encuesta fue enviada a los 403 participantes y se obtuvo una tasa de respuesta cercana al 55%, generando una muestra de 222 observaciones.

Para determinar si la muestra era representativa de la población se elaboraron *clusters* agrupando los registros en base al ranking de entrada. Una observación pertenece al primer *cluster* si en su generación fue clasificada entre las primeras diez mejores oportunidades, una observación pertenece al segundo *cluster* si en su generación fue clasificada entre las segundas diez mejores oportunidades y así sucesivamente. Dado que la muestra corresponde a un 55% de la población, se espera que los *clusters* de la muestra contengan aproximadamente un 55% de las observaciones que tienen los *clusters* de la población. La ilustración 1 muestra los resultados obtenidos luego del *clustering* donde se puede observar que la muestra pareciera ser representativa.

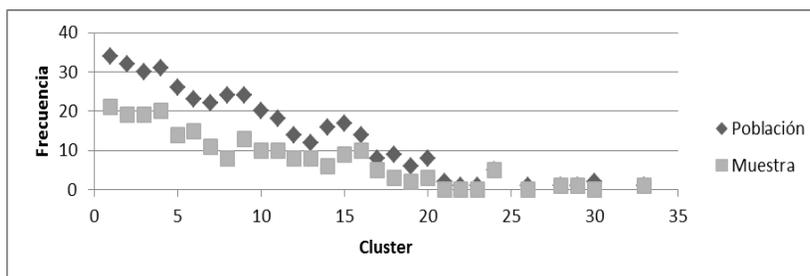


Figura 1: Clustering del Ranking de Entrada

Para confirmar estos resultados se utilizó el test de Kolmogorov-Smirnov, el cual se encuentra en la Tabla 1. Como se puede observar, no existe suficiente evidencia para señalar que las distribuciones del ranking de la población y la muestra son diferentes.

Tabla 1: Resultados test Kolmogorov-Smirnov

| Smaller Group | D | P-Value | Exact |
|---------------------|---------|---------|-------|
| 0 | 0,0182 | 0,939 | - |
| 1 | -0,1052 | 0,120 | - |
| Combined K-S | 0,1052 | 0,240 | 0,220 |

Otro aspecto considerado corresponde al porcentaje de éxito del programa, el cual asciende a un 15% (57 *startups*), mientras que en la muestra este asciende a un 17% (38 *startups*). Por lo que la muestra también es representativa en este aspecto.

5. METODOLOGÍA

En base al análisis de la literatura se llegó a la selección de cuatro atributos que estarían relacionados con el descubrimiento de mejores oportunidades, un mejor desempeño y una mayor probabilidad de sobrevivencia a través del tiempo. Luego se analizó estadísticamente cada una de las reglas, para posteriormente utilizando árboles de decisión, combinar las reglas estableciendo tanto reglas compensatorias como también reglas no compensatorias. Finalmente una vez construidos los modelos se compara su desempeño con la predicción entregada por el ranking de la consultora.

5.1. Selección de reglas

La literatura destaca la importancia que tiene la reactividad a los cambios, donde los mejores competidores, los más exitosos son aquellos que saben moverse y mantenerse al filo del tiempo (Stalk, 1988).

A nivel industrial es necesario ser capaz de responder de manera rápida y competitiva al fuerte ritmo de cambios tecnológicos, gatillado por periodos de obsolescencia más rápidos y la necesidad de adoptar riesgo para entregar respuestas oportunas (Bird & Page West, 1998).

Este ritmo de cambio en las tecnologías de producción, en los procesos industriales y en los productos presentes en el mercado ha incrementado la volatilidad. Para poder adaptarse a este nuevo ambiente las empresas deben innovar con mayor frecuencia y ser capaces de implementar las innovaciones lo más rápido posible. La literatura describe este fenómeno como competición basada en el tiempo (Askenazy, Thesmar, & Thoenig, 2006).

Jaffe & Trajtenberg (1996) estudian el flujo temporal de cita de patentes y lo describen como un proceso con forma de u invertida compuesto por la difusión y la obsolescencia. El descubrimiento de oportunidades debería seguir un patrón similar, siendo exitosas las empresas que explotan las oportunidades del mañana y fracasando aquellas que explotan las oportunidades del pasado. En esta visión el tiempo correspondería a un costo de oportunidad, al cual se asocia el riesgo de ser superado por los competidores (Askenazy et al., 2006).

En conjunto los argumentos previamente presentados señalan que debido al impacto incremental que tiene la obsolescencia tecnológica, de procesos y de productos en la competitividad de las empresas, estas se ven forzadas a incorporar el factor temporal a su proceso de innovación o a enfrentar el riesgo de que la oportunidad sea explorada por sus competidores. En línea con lo anterior se propone la primera regla:

R1: La idea debe ser factible como máximo desde hace 2 años

Uno de los principales indicadores de factibilidad es la construcción de un prototipo. El enfoque de *Lean Startup* define al Producto Mínimo Viable (MVP) como una versión básica del producto, la cual al ser testeada entregara retroalimentación acerca de los usuarios y el uso que le dan al producto, pudiendo de esta manera visualizar las funcionalidades faltantes para lograr resolver una necesidad real (Blank, 2007; Björk, Ljungblad, & Bosch, 2013). Ries (2011) establece que la construcción del MVP puede utilizarse para determinar si los consumidores están dispuestos a pagar por el producto y verificar si la solución es factible en el horizonte de tiempo planteado.

Audretsch, Bönte, & Mahagaonkar (2012) definen la construcción de un prototipo como parte del proceso de comercialización del producto final, ya que este

facilitaría el proceso de patentes. Por otra parte, es factible determinar los procesos necesarios para la producción a gran escala, los recursos necesarios e ir estableciendo relaciones con eventuales proveedores. También destacan que la construcción de un prototipo disminuye la incertidumbre propia del proyecto, le entrega un valor agregado a las patentes y facilitaría el acceso a capital de riesgo (Audretsch et al., 2012).

Por otra parte, el enfoque de prototipos permite adelantar las etapas de validación de hipótesis y provee una mayor comprensión del ambiente al facilitar la retroalimentación y una mayor flexibilidad a lo largo de los proyectos (Hilal & Soltan, 1992).

Los argumentos presentados previamente sugieren que la construcción de un prototipo impacta de manera positiva en el desarrollo de un *startup*, ya que permite obtener *feedback* y validar hipótesis con el usuario final, asegurar la factibilidad técnica, identificar procesos, recursos y proveedores, además de facilitar el acceso a capital de riesgo. En consecuencia se propone la siguiente regla:

R2: El equipo debe haber desarrollado un prototipo.

La literatura analiza la relación entre el emprendimiento y la incertidumbre, principalmente desde dos perspectivas (1) el descubrimiento de oportunidades y (2) aspectos psicológicos propios del emprendedor.

Comenzar un emprendimiento basado en un nuevo producto o servicio, determinar la aceptación que tendrá un producto en el mercado, eventuales cambios en la demanda, en las tecnologías que utilizan las industrias o en las relaciones estratégicas entre los competidores, son escenarios que ofrecen a los emprendedores un potencial para la innovación. El emprendedor debe convivir con la incertidumbre, tomando

decisiones para las cuales prácticamente no existe información (Russell, 1992; Busenitz & Barney, 1994).

De acuerdo a esto, compañías situadas en ambientes de mayor incertidumbre deberían implementar un mayor número de innovaciones que las compañías de ambientes menos dinámicos (Russell, 1992; Davis, Eisenhardt, Bingham, 2009).

Leatherbee (2012) establece que la existencia de incertidumbre y abundancia de recursos son condiciones relevantes para la existencia de oportunidades, ya que permiten quebrantar el *status quo* y reasignar los recursos hacia nuevas oportunidades.

En cuanto a los aspectos psicológicos, McMullen & Shepherd (2006) destacan una corriente literaria la cual distingue entre quienes son capaces de emprender pese a la incertidumbre y quienes no, estableciendo diferencias en términos de motivación, actitud y propensión al riesgo.

En consecuencia, las compañías situadas en ambientes de incertidumbre son capaces de descubrir y explotar nuevas y mejores oportunidades que aquellas compañías de ambientes menos dinámicos. Además, quienes son capaces de explotar oportunidades en ambientes de incertidumbre contarían con ciertos aspectos psicológicos favorables al momento de emprender, por lo que de acuerdo a estos argumentos se propone la tercera regla:

R3: El equipo debe reconocer la incertidumbre

El conocimiento es una ventaja competitiva, es la fuente del crecimiento y desarrollo. Diversos estudios han demostrado relaciones entre el nivel educacional y el emprendimiento.

Una de las etapas claves del emprendimiento corresponde al descubrimiento de oportunidades, donde Shane (2000; pp. 467) establece “*Potential entrepreneurs should look to discover opportunities in what they know...*”

También es posible relacionar el nivel educacional con la adquisición de habilidades fundamentales para el emprendimiento como la comunicación, la toma de decisiones, la resolución de problemas, habilidades de trabajo en equipo y la capacidad de adaptarse a ambientes de cambios constantes (Cooper, Gimeno-Gascón, & Woo, 1997; Davidsson & Honig, 2003; Magoutas, Papadogonas, & Sfakianakis, 2012).

Por último, la literatura se ha concentrado en relacionar el impacto que tiene el capital humano en el desempeño de la firma desde dos perspectivas: (1) Supervivencia y (2) Crecimiento.

Si bien esperaríamos que compañías administradas por personas con mayor educación tuvieran una mayor tasa de supervivencia, y por ende este fuera un factor que afectara positivamente el acceso a capital de riesgo, Kangasharju & Pekkala (2002) demuestran que esto es cierto solo en períodos de recesión económica y que el nivel educacional tiene un efecto contrario en períodos de auge económica. Este aspecto se explica ya que en periodos de auge económico, los profesionales deberían realizar un *trade-off* entre las ofertas que les entrega el mercado y su deseo de emprender.

Los *startups* conformados por personas con mayor educación tendrían un mayor potencial de crecimiento (A. C. Cooper, 1993), debido a que les permitiría tener una mayor comprensión del mercado y por ende explotarían la demanda de una mejor manera (Kangasharju & Pekkala, 2002).

En síntesis, un mayor nivel educacional se relaciona con el descubrimiento de mejores oportunidades, con la adquisición de ciertas habilidades blandas que facilitan el emprendimiento, con una mayor tasa de crecimiento y en algunos casos, con una mayor probabilidad de sobrevivencia. En base a estos argumentos se propone la cuarta regla:

R4: Todos los miembros del equipo deben haber finalizado la universidad

5.2. Análisis Estadístico

Primero se utilizaron tablas para observar y analizar el potencial de predicción de cada una de las reglas. Luego se elaboraron cuatro modelos logit de la forma:

$$Demoday = f(\beta_0 + \beta_1 * Regla_i) \quad (1)$$

Este análisis tiene dos objetivos: (1) determinar si el efecto de cada una de las reglas es consistente con lo descrito por la literatura y (2) utilizar el test t para determinar el nivel de significancia de cada una de las reglas respecto a la probabilidad de selección.

Posteriormente se elaboró un modelo logit que relacionara el Demoday con las cuatro reglas en simultaneo, para poder determinar el efecto conjunto que tendrían y evaluar nuevamente el nivel de significancia. El modelo completo tiene la siguiente forma:

$$Demoday = f(\beta_0 + \beta_1 * Regla_1 + \beta_2 * Regla_2 + \beta_3 * Regla_3 + \beta_4 * Regla_4) \quad (2)$$

5.3. Combinación de reglas

La dificultad de combinar las reglas correctamente radica en: (1) determinar cuáles son compensatorias y cuáles son no compensatorias y (2) el hecho de que utilizar reglas no compensatorias implica que el orden en que se sitúan las reglas tiene un

impacto fundamental, ya que una regla se revisará solo si las reglas previas han sido aprobadas (Grzech, 2009). Una metodología que permite enfrentar estas dificultades, explotando la información disponible corresponde a los árboles de decisión.

Los árboles de decisión son una técnica de aprendizaje supervisado, es decir, utilizando un set de registros pre-clasificados conocido como set de entrenamiento generan un árbol de clasificación. El set de entrenamiento le permite al modelo encontrar una estructura apropiada para el árbol, es decir, explorar el espacio de hipótesis en busca de un clasificador.

El árbol contiene nodos y ramas permitiendo visualizar la secuencia de evaluaciones que permiten obtener una clasificación. Cada nodo representa un atributo evaluado, donde se realizan pruebas y las ramas resultantes representan el resultado de estas pruebas. Por lo que para poder clasificar un registro es necesario recorrer el árbol desde el nodo raíz (primer atributo evaluado) hasta llegar a un nodo hoja, es decir, un nodo que entrega una clasificación donde el camino recorrido está determinado por los valores del registro (Espejo, Romero, Hervás, & Ventura, 2005).

Los árboles construyen un modelo de la regularidad existente en los datos. Además de clasificar, son capaces de extraer una estructura que representa, en cierta medida, el patrón de comportamiento que hay asociado a la muestra sobre la que se ha inducido (Sierra Araujo, 2006). El principal supuesto de estos modelos es el principio inductivo, es decir, si el algoritmo es capaz de tener un buen performance en el set de entrenamiento, en la medida en que éste sea representativo de la población, el modelo tendrá un buen desempeño al clasificar nuevas instancias.

Además, el problema de clasificación asume que las clases son disjuntas, es decir, una instancia sólo pertenece a una clase. Debido a esto un árbol de clasificación conducirá hasta una y solo una hoja y por lo tanto a la instancia se le asignará una sola clase.

5.3.1. Construcción del árbol

El algoritmo de construcción del árbol es un proceso iterativo, y en su forma más básica, funciona de la siguiente manera:

1. En caso de que todos los registros pertenezcan a la misma clase, el algoritmo crea una hoja identificando a los registros con la clase respectiva.
2. En caso de que todos los registros tengan el mismo valor para **todos** los atributos, el algoritmo crea una hoja identificando a los registros con la clase más común.
3. En caso contrario, el algoritmo selecciona al atributo que mejor separa los registros de las distintas clases y utiliza este registro como nodo raíz. Luego, divide el set de entrenamiento de acuerdo a este atributo y para cada rama resultante continua la construcción del árbol de manera recursiva (Punto 1) hasta que todas las ramas finalicen en una hoja.

Para la creación del modelo se optó por utilizar *cross-validation* debido a que no se dispone de una gran cantidad de datos. Esta estrategia realiza tantas evaluaciones como indica el parámetro *fold*s. Para este trabajo se utilizan 10 *fold*s, es decir, se dividen los registros en 10 grupos y en cada evaluación se toman las instancias de cada grupo

como datos de evaluación y el resto como set de entrenamiento para construir el modelo. Los errores calculados corresponden al promedio de todas las ejecuciones.

De manera natural, los inversionistas ponderan de manera diferente el cometer un Error Tipo I, descartar una oportunidad incorrectamente, y el cometer un error de Tipo II, seleccionar una oportunidad correctamente. Por lo que es necesario incorporar las preferencias del evaluador, penalizando más el Error Tipo I, en caso de no querer descartar potenciales oportunidades en una primera etapa o penalizando más el Error Tipo II, en la medida en que se incrementa la aversión al riesgo.

Para la construcción de los Modelos se utilizó la matriz de costos que se presenta en la Tabla 8. Es posible observar que se penaliza más el descartar potenciales oportunidades incorrectamente, que el seleccionar oportunidades incorrectas, lo cual es propio de la etapa de preselección de oportunidades.

Tabla 2: Matriz de Costos

| | No Aprueba | Aprueba |
|------------------------|-------------------|----------------|
| No Seleccionado | 0 | 1 |
| Seleccionado | 3 | 0 |

Los valores que se ingresan a la matriz debiesen ser capaces de representar el costo de no seleccionar una oportunidad exitosa, es decir, los potenciales ingresos que podría dejar de recibir un *Start-up* exitoso que no fue apoyado, y los costos de seleccionar una oportunidad no exitosa, los cuales corresponden básicamente a los US\$40.000 que entrega el programa.

Además, en caso de que el algoritmo utilice iguales pesos se genera un problema común que enfrentan estos algoritmos conocido como *unbalanced data*. Esta

problemática se debe a la baja tasa de éxito del programa, la cual bordea un 15% y produce que el algoritmo clasifique a todos los registros como “No seleccionados”. Además de las virtudes mencionadas previamente de la incorporación de la matriz de costos, esta solución permite resolver el desbalance de datos (Chawla, 2005; Sun, Kamel, Wong, & Wang, 2007; He & Garcia, 2009).

Finalmente, una vez construidos los modelos el algoritmo de árboles de decisión ofrece “herramientas de poda”, cuyo objetivo es eliminar aquellas ramas que se sobre ajustan al set de entrenamiento y limitan la capacidad de generalizar de los modelos obtenidos.

6. RESULTADOS

6.1. Análisis Individual

6.1.1. Factibilidad

Los resultados corresponden a la respuesta entregada por los participantes a la pregunta Q1 del cuestionario (Anexo A), aprueban la regla quienes contestaron las alternativas “b”, “c” o “d”. El resto de las respuestas desaprobaban la primera regla. La Tabla 3 resume las respuestas entregadas por los encuestados.

Tabla 3: Respuesta de los participantes a la pregunta Q1

| Respuesta | Frecuencia |
|--------------------------|--------------|
| It is still not feasible | 30 (13,51 %) |
| Less than a year ago | 89 (40,09 %) |
| One year ago | 37 (16,67 %) |
| Two years ago | 34 (15,32 %) |
| Three years ago | 22 (9,91 %) |
| Four years ago | 2 (0,90 %) |
| Five or more years ago | 8 (3,60%) |

La Tabla 4 contrasta la aprobación de la primera regla contra la selección para el Demoday. Es posible observar que 160 (72 %) *startups* consideran que es factible desarrollar su idea desde hace no más de dos años.

Tabla 4: Aprobación primera regla y selección Demoday

| Demoday / R1 | No Aprueba | Aprueba | Total |
|-----------------|--------------|---------------|---------------|
| No Seleccionado | 56 (25,22 %) | 128 (57,65 %) | 184 (82,88 %) |
| Seleccionado | 6 (2,70 %) | 32 (14,41 %) | 38 (17,11 %) |
| Total | 62 (27,92 %) | 160 (72,06 %) | 222 |

De los 38 seleccionados para presentar su emprendimiento en el Demoday con los que cuenta la muestra, solo 6 (15,7 %) consideran que su idea es factible desde hace más de dos años.

6.1.2. Prototipo

Los resultados corresponden a la respuesta entregada por los participantes a la pregunta “What phase of development is your Company in?” del cuestionario completado al momento de postular. Aprueban quienes respondieron “Scaling Sales” o “Functional Product with Users”. Desaprueban la segunda regla quienes respondieron “Concept” o “Working Prototype in Development”. La Tabla 5 resume las respuestas entregadas por los encuestados.

Tabla 5: Respuesta de los participantes a la pregunta "Phase of Development"

| Respuesta | Frecuencia |
|----------------------------------|-------------------|
| Concept | 23 (10,36 %) |
| Functional Product with Users | 76 (34,23 %) |
| Scaling Sales | 10 (4,50 %) |
| Working Prototype in Development | 113 (50,90%) |

Como es posible observar en la Tabla 6 solo 86 (38,73 %) *startups* postulan habiendo desarrollado un prototipo. Sin embargo, de los 136 que no aprueban la regla 113 (50,9 %) estarían desarrollando su prototipo al momento de postular.

Tabla 6: Aprobación segunda regla y selección Demoday

| Demoday / R2 | No Aprueba | Aprueba | Total |
|------------------------|-------------------|----------------|---------------|
| No Seleccionado | 116 (52,25 %) | 68 (30,63 %) | 184 (82,88 %) |
| Seleccionado | 20 (9 %) | 18 (8,1 %) | 38 (17,11 %) |
| Total | 136 (61,25 %) | 86 (38,73%) | 222 |

Al momento de analizar a los seleccionados para presentarse en el Demoday 18 aprueban la regla, es decir, ya habrían desarrollado su prototipo. En cuanto a los 20 *startups* que desaprueban la regla, 17 estarían actualmente desarrollando su prototipo.

6.1.3. Incertidumbre

Los resultados corresponden a la respuesta entregada por los participantes a la pregunta Q10 del cuestionario (Anexo A), aprueban la regla quienes contestaron las alternativas “a” o “b”. El resto de las respuestas desaprueban la tercera regla. La Tabla 7 resume las respuestas entregadas por los encuestados.

Tabla 7: Respuesta de los participantes a la pregunta Q10

| Respuesta | Frecuencia |
|--|-------------------|
| The main uncertainty is the market | 99 (44,59 %) |
| The main uncertainty is the product or service | 63 (28,37 %) |
| The startup face no uncertainties | 60 (27,02 %) |

La Tabla 8 contrasta la tercera regla contra la participación en el Demoday. De los 162 (72,97 %) participantes que reconocen incertidumbre, 99 la asocian al tamaño del mercado y los 63 restantes la asocian al desarrollo del producto o servicio.

Tabla 8: Aprobación tercera regla y selección Demoday

| Demoday / R3 | No Aprueba | Aprueba | Total |
|------------------------|-------------------|----------------|---------------|
| No Seleccionado | 55 (24,77 %) | 129 (58,1 %) | 184 (82,88 %) |
| Seleccionado | 5 (2,25 %) | 33 (14,86 %) | 38 (17,11 %) |
| Total | 60 (27,02 %) | 162 (72,96%) | 222 |

De los 38 seleccionados para el Demoday solo 5 no aprueban la regla, en otras palabras, creen que su *startups* no enfrenta incertidumbres.

6.1.4. Nivel educacional

Los resultados corresponden a la respuesta entregada por los participantes a la pregunta “What is the highest level of education you had completed?” del cuestionario completado al momento de postular. Aprueban los grupos cuya totalidad de participantes respondió alguna de las siguientes opciones “College Degree (Professional Degree)”, “Masters” o “Ph.D.”. Desaprueban la cuarta regla quienes respondieron “Less than High School”, “High School”, “2-Year College Degree (Minor)” o “4-Year College Degree (Major)”. La Tabla 9 resume las respuestas entregadas por los encuestados.

Tabla 9: Respuesta de los participantes a la pregunta "What is your Highest Educational Level"

| Respuesta | Frecuencia |
|--------------------------------------|-------------------|
| Less than High School | 1 (0,45 %) |
| High School | 10 (4,50 %) |
| 2-Year College Degree (Minor) | 6 (2,70 %) |
| 4-Year College Degree (Major) | 101 (45,49 %) |
| College Degree (Professional Degree) | 44 (19,81 %) |
| Masters | 54 (24,32 %) |
| Ph.D | 6 (2,70 %) |

La Tabla 10 exhibe los resultados al contrastar la cuarta regla con la participación en el Demoday, donde 104 (46,84 %) equipos han finalizado su educación universitaria. Sin embargo, en 67 de los 118 equipos que no aprueban la regla hay al menos un integrante que si finalizo su educación universitaria.

Tabla 10: Aprobación cuarta regla y selección Demoday

| Demoday / R4 | No Aprueba | Aprueba | Total |
|------------------------|-------------------|----------------|---------------|
| No Seleccionado | 100 (45,04 %) | 84 (37,83 %) | 184 (82,88 %) |
| Seleccionado | 18 (8,1 %) | 20 (9 %) | 38 (17,11 %) |
| Total | 118 (53,15 %) | 104 (46,84 %) | 222 |

Como muestra la Tabla 10, solo 20 del total de equipos seleccionados para el Demoday aprueban la regla, pero en 30 de los 38 equipos es posible encontrar al menos a una persona que haya finalizado su educación universitaria.

6.2. Análisis Estadístico

La Tabla 11 presenta los resultados obtenidos para los modelos presentados en la ecuación (1). De las cuatro reglas simples solo serían significativas la tercera y cuarta, correspondientes al reconocimiento de la incertidumbre y el nivel educacional. La primera regla, es decir, el tiempo desde que es factible el desarrollo de la idea, sería significativo al 7%. Además, las tres reglas mencionadas tendrían un efecto positivo en la probabilidad de selección, siendo coherentes con lo establecido en la literatura.

Tabla 11: Modelos individuales

| | Constraint | Coefficient | P-Value | Chi-2 |
|--------------------|------------|-------------|---------|--------|
| Feasibility Time | -2,233 | 0,847 | 0,073 | 0,0550 |
| Prototype | -1,757 | 0,428 | 0,2323 | 0,234 |
| Uncertainty | -2,397 | 1,034 | 0,041 | 0,024 |
| Level of Education | -2,14 | 0,675 | 0,022 | 0,020 |

En la Tabla 12 se presentan los resultados obtenidos para el modelo completo descrito en la ecuación (2). El modelo tiene un valor-p de 0,0015 por lo que es significativo. En esta ocasión la primera y tercera regla son significativas (*valor - p < 0.05*). La segunda regla, correspondiente al desarrollo de un prototipo, sería significativa en torno a un 13% y la cuarta regla, es decir, el nivel educacional, deja de ser significativa. Además, nuevamente la aprobación de cada una de las reglas impacta de manera positiva en la selección para el Demoday, en línea con lo establecido en la literatura.

Tabla 12: Modelo completo

| | Feasibility Time | Prototype | Uncertainty | Level of Education |
|-------------|------------------|-----------|-------------|--------------------|
| Coefficient | 0,989 | 0,572 | 1,264 | 0,420 |
| P-Value | 0,040 | 0,13 | 0,015 | 0,256 |

6.3. Modelo 1

El árbol resultante se muestra en la Figura 2 y tiene como reglas no compensatorias a la primera y tercera regla, es decir, deben cumplirse ambas para ser seleccionado y como regla compensatoria el cumplimiento de la segunda o de la cuarta regla.

En el nodo raíz tenemos la totalidad seleccionados para el Demoday (38) y la totalidad de la muestra (222 observaciones). El algoritmo seleccionó como nodo raíz a la tercera regla, es decir, el reconocimiento de la incertidumbre. Quienes desapruban la regla toman la rama izquierda y quienes aprueban toman la rama derecha.

La rama izquierda el algoritmo comprende un total de 60 observaciones, donde 55 (92 %) *startups* no fueron seleccionados para el Demoday. Debido a la homogeneidad que presentan estos registros el algoritmo clasifica a quienes no aprueban como “Rechazado”, clasificando correctamente a 55 *startups* y de manera errónea las 5 observaciones restantes. Por otra parte la rama derecha abarca un total de 162 observaciones, donde el 20 % fue seleccionado para el Demoday y el 80 % restante no fue seleccionado, en otras palabras, los registros presentes en esta clasificación presentan una mayor heterogeneidad. Debido a esta característica, el algoritmo no es capaz de entregar una clasificación y decide continuar evaluando y selecciona como atributo la

primera regla. El proceso se repite hasta que cada una de las ramas entregue una clasificación para los registros.

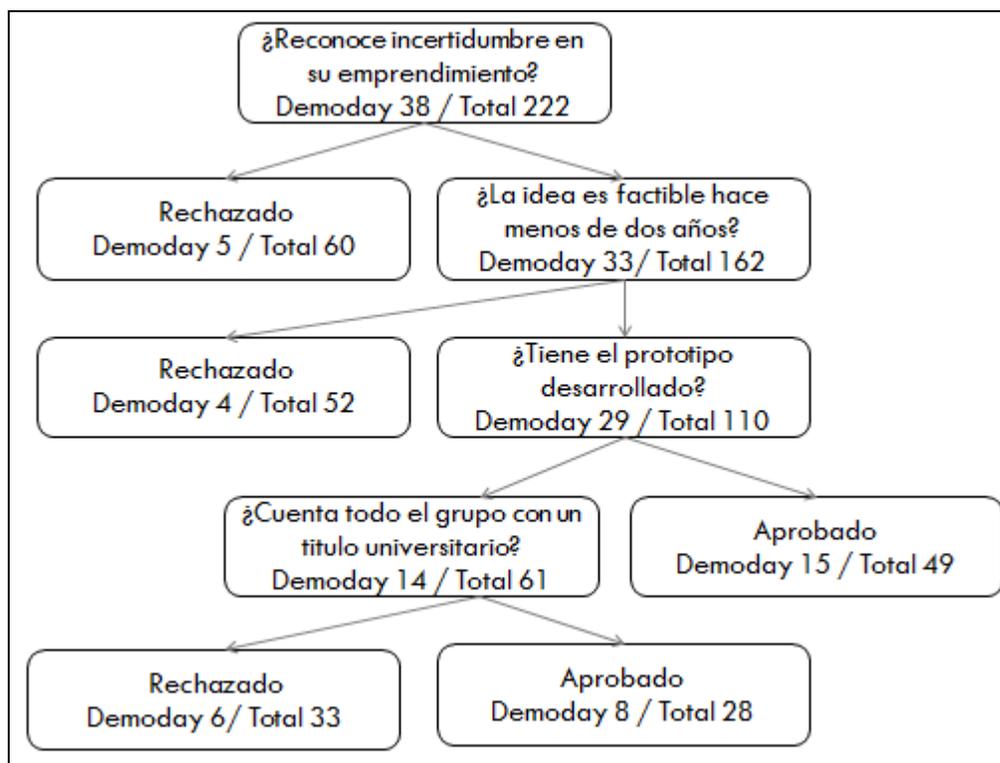


Figura 2: Clasificación mediante un árbol de decisión Modelo 1

Los resultados de la clasificación se presentan en la Tabla 13. El árbol clasifica correctamente a 153 (68,91 %) de los *startups*, correspondientes a quienes no aprueban la combinación de reglas y no son seleccionados (130 *startups*) y a quienes aprueban y son seleccionados (23 *startups*).

Tabla 13: Clasificación del árbol de decisión Modelo 1

| | No Aprueba | Aprueba |
|-----------------|---------------|--------------|
| No Seleccionado | 130 (58,55 %) | 54 (24,32 %) |
| Seleccionado | 15 (6,75 %) | 23 (10,36 %) |

El modelo seleccionó 77 *startups*, por lo que los jueces deben seleccionar igual cantidad de oportunidades de negocio. Considerando los 77 proyectos con mejor evaluación de los jueces, solo 17 fueron seleccionados para el Demoday, por lo que la clasificación utilizando árboles de decisión representa un incremento del 35% en la efectividad de selección (Ver Tabla 14).

Tabla 14: Clasificación de los jueces Modelo 1

| | No Aprueba | Aprueba |
|-----------------|---------------|--------------|
| No Seleccionado | 124 (55,85 %) | 60 (27,02 %) |
| Seleccionado | 21 (9,45 %) | 17 (7,65 %) |

Al momento de analizar la significancia estadística de ambas evaluaciones, la combinación de reglas es significativa (*Valor - p = 0,00*), mientras que el ranking de los jueces no sería significativo (*Valor - p ~ 0,16*).

Dado que el árbol presentado en la Figura 2 involucra reglas compensatorias, se forzó al algoritmo a crear dos nuevos árboles utilizando solo reglas no compensatorias. Para esto se le prohibió utilizar en primera instancia la segunda regla, correspondiente al desarrollo de un prototipo y luego la cuarta regla, es decir, el nivel educacional. Dando origen respectivamente a los Modelos 2 y 3 respectivamente.

6.4. Modelo 2

La Figura 3 muestra el árbol resultante al utilizar solo la primera, tercera y cuarta reglas. El objetivo de obtener un modelo que combinara las reglas de manera no compensatoria se cumple.

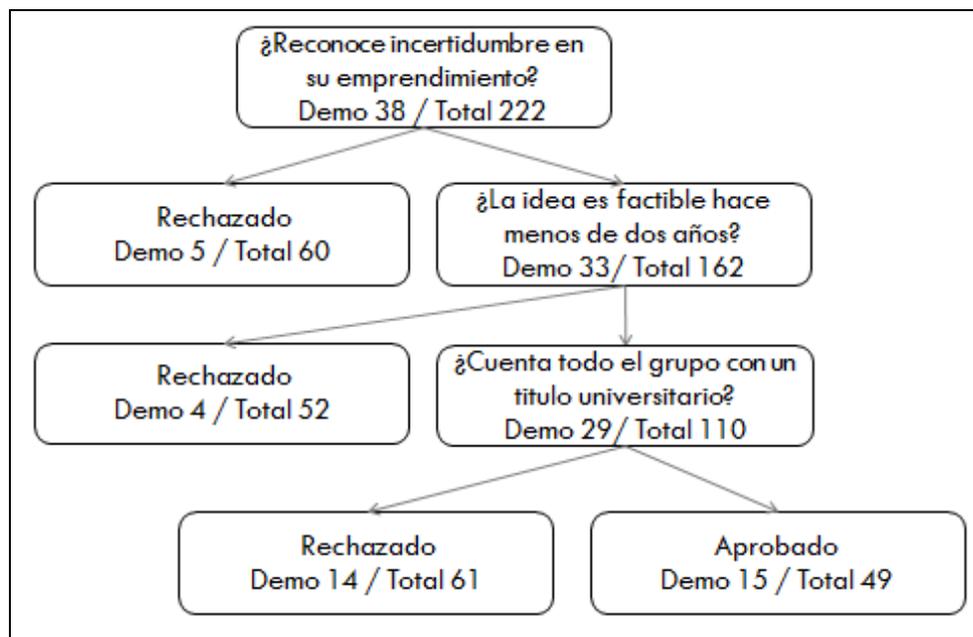


Figura 3: Clasificación mediante un árbol de decisión Modelo 2

Los resultados de la clasificación se muestran en la Tabla 15. El modelo es capaz de clasificar correctamente a 165 (74,32 %) de los *startups*, si bien esto representa un incremento respecto al Modelo 1, es capaz de clasificar correctamente solo 15 seleccionados para el Demoday, en otras palabras, pierde eficacia al detectar a los seleccionados respecto al Modelo 1.

Tabla 15: Clasificación del árbol de decisión Modelo 2

| | No Aprueba | Aprueba |
|------------------------|---------------|--------------|
| No Seleccionado | 150 (67,56 %) | 34 (15,31 %) |
| Seleccionado | 23 (10,36 %) | 15 (6,75 %) |

Respecto a la comparación de los jueces, nuevamente el modelo demuestra tener una mayor efectividad. Como muestra la Tabla 16, los jueces clasifican correctamente el 70,72 % de las oportunidades. Dado que el modelo selecciona 49 *startups*, al considerar las mejores 49 evaluaciones de los jueces solo se detectan 11 seleccionados para el

Demoday. Por lo que el modelo representaría un incremento de un 36,36 % en la efectividad de la selección.

Tabla 16: Clasificación de los jueces Modelo 2

| | No Aprueba | Aprueba |
|-----------------|---------------|--------------|
| No Seleccionado | 146 (65,76 %) | 38 (17,11 %) |
| Seleccionado | 27 (12,16 %) | 11 (4,95%) |

Nuevamente la combinación de reglas resulta significativa ($Valor - p \sim 0,00$) y el ranking entregado por los jueces para las mejores 49 evaluaciones no es significativo ($Valor - p \sim 0,26$).

6.5. Modelo 3

Por último, el Modelo 3 combina la primera, segunda y tercera regla como muestra la Figura 4. Nuevamente se cumple el objetivo de utilizar las reglas solo de manera no compensatoria.

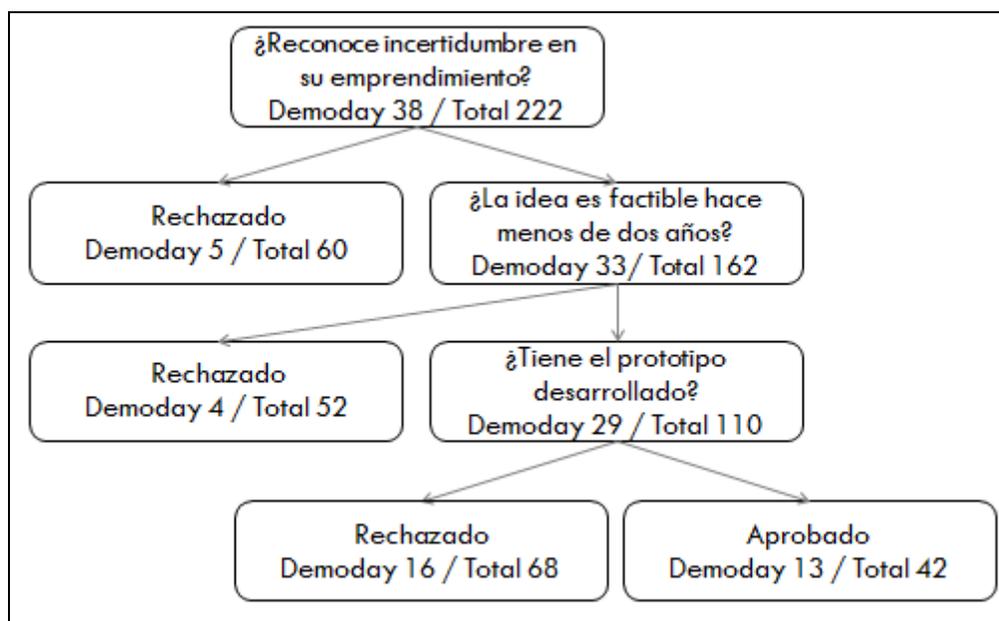


Figura 4: Clasificación mediante un árbol de decisión Modelo 3

La Tabla 17 resume la clasificación efectuada por el Modelo 3, el cual clasifica correctamente a 168 (75,67 %) de los registros y selecciona correctamente a 13 *startups* para el Demoday. Al igual que en el segundo modelo se observa una disminución de la eficiencia de selección respecto al Modelo 1.

Tabla 17: Clasificación del árbol de decisión Modelo 3

| | No Aprueba | Aprueba |
|-----------------|--------------|-------------|
| No Seleccionado | 155 (69,81%) | 29 (15,78%) |
| Seleccionado | 25 (11,26 %) | 13 (5,85 %) |

La Tabla 14 presenta los resultados de la clasificación de los jueces, quienes clasifican correctamente a 158 (71,17 %) de los participantes, pero clasifican correctamente a solo 8 *startups* para el Demoday.

Tabla 18: Clasificación de los jueces Modelo 3

| | No Aprueba | Aprueba |
|-----------------|---------------|--------------|
| No Seleccionado | 150 (67,56 %) | 34 (15,31 %) |
| Seleccionado | 30 (13,51 %) | 8 (3,6 %) |

En síntesis, al momento de detectar oportunidades la clasificación utilizando el árbol identifica 13 oportunidades y es significativo ($Valor - p = 0,01$), mientras que los jueces detectan solo 8 y su evaluación para los mejores 42 postulantes no es significativa ($Valor - p = 0,71$). Por lo que el Modelo 3 representa un incremento de un 62, 5% en la efectividad al seleccionar *startups*.

7. VALIDACIÓN DE LOS MODELOS

Debido a que los modelos utilizan la totalidad de los datos para su construcción, se utilizó a los participantes de la Generación 8 de Start-Up Chile como *out sample*. En esta generación se seleccionaron un total de 85 participantes para el programa, de los cuales 16 fueron seleccionados para presentarse en el Demoday.

El Modelo 1 seleccionó un total de 25 *startups* y 7 de estos fueron seleccionados para participar en el Demoday, mientras que al seleccionar los 25 participantes con mejor ranking solo encontramos 5 participantes para el Demoday. En esta ocasión, tanto el Modelo 1 como las mejores 25 evaluaciones de los jueces no serían significativos teniendo Valores-P iguales a 0,168 y 0,858 respectivamente.

En cuanto al Modelo 2, este tiene un peor desempeño que la evaluación de los jueces ya que selecciona un total de 17 *startups*, detectando solo 3 participantes para el Demoday. Al considerar los 17 mejor evaluados por los jueces encontramos 4 participantes para el Demoday. Además, si bien ninguna de las evaluaciones es significativa (Valores-P iguales a 0,89 y 0,58 respectivamente), la evaluación de los jueces sería más significativa que los resultados entregados por el modelo.

Finalmente, el Modelo 3 selecciona un total de 16 oportunidades detectando 6 participantes para el Demoday y al considerar los mejores 16 evaluados por los jueces se detectan solo 4 oportunidades. A diferencia de los casos anteriores, el Modelo 3 resulta significativo (*Valor - P = 0,04*), pero la evaluación de los jueces se mantiene como una variable no significativa (*Valor - P = 0,48*).

8. CONCLUSIONES

El principal objetivo de este trabajo era desarrollar un *framework* que fuera capaz de clasificar oportunidades de negocio utilizando reglas simples como criterio de decisión. Para lograr combinar las reglas de manera efectiva se utilizaron árboles de decisión, ya que estos utilizan la información disponible para identificar los atributos que discriminan de mejor manera y permiten obtener un modelo que clasifica utilizando tanto reglas compensatorias como no compensatorias.

Utilizando datos de un programa de gobierno fue posible construir los árboles y evaluar su eficacia al momento de clasificar, contrastando este resultado contra la evaluación de expertos. Las combinaciones de reglas provenientes de los árboles demostraron ser estadísticamente significativas y además tener una mayor eficiencia que la evaluación experta basada en los enfoques estratégicos de posicionamiento y recursos, al momento de predecir el éxito.

La principal debilidad del modelo es propia de los métodos de clasificación y corresponde a que una vez clasificadas las oportunidades no existiría un criterio de selección, en otras palabras, dentro de una misma clasificación el modelo no es capaz de justificar la selección de una oportunidad por sobre otra.

Además, este trabajo confirma el impacto positivo que tienen para los *startups* una idea novedosa, el desarrollo de un prototipo, la incertidumbre y un mayor nivel educacional en el acceso a capital de riesgo.

Finalmente, el uso de esta metodología les permitiría tanto a empresas de capital de riesgo como a programas de gobierno asociados al emprendimiento hacer un uso

efectivo de la información con la que cuentan, siendo capaces de comprender y revisar de manera objetiva las reglas que utilizan para seleccionar oportunidades de negocio.

La continuidad de la investigación debe centrarse en los siguientes tres puntos: (1) Deben evaluarse nuevas reglas simples como la experiencia previa del equipo en *startups*, el tipo de cambio que origino la oportunidad, el conocimiento del mercado y otros aspectos propios de la industria; (2) Se deben evaluar otras variables de performance como la tasa de sobrevivencia, el crecimiento de los *startups* y el volumen de ventas por nombrar algunas; (3) Finalmente se puede explotar de mejor manera el algoritmo de árboles de decisión, ya que estos tienen la capacidad de determinar las reglas simples por sí mismos.

BIBLIOGRAFÍA

- Askenazy, P., Thesmar, D., & Thoenig, M. (2006). On the relation between organisation practices and new technologies: the role of (time based) competition. *The Economic Journal*, *116*, 128–154.
- Audretsch, D. B., Bönte, W., & Mahagaonkar, P. (2012). Financial signaling by innovative nascent ventures: The relevance of patents and prototypes. *Research Policy*, *41*(8), 1407–1421.
- Bingham, C. B., Eisenhardt, K. M., & Furr, N. R. (2011). Which Strategy When? *MIT Sloan Management Review*, *53*(1), 71–77.
- Bird, B. J., & Page West, G. (1998). Time and Entrepreneurship. *Entrepreneurship Theory and Practice*.
- Björk, J., Ljungblad, J., & Bosch, J. (2013). *Lean Product Development in Early Stage Startups*.
- Blank, S. G. (2007). *The Four Steps to the Epiphany*. Cafepress. com (p. 281).
- Busenitz, L. W., & Barney, J. B. (1994). Biases and heuristics in strategic decision making: Differences between entrepreneurs and managers in large organizations. *Academy of Management Best Papers Proceedings*.
- Chawla, N. V. N. (2005). Data Mining for Imbalanced Datasets- An Overview. *Data Mining and Knowledge Discovery Handbook*, 853–867. doi:10.1007/0-387-25465-X_40
- Cooper, A. C. (1993). Challenges in predicting new firm performance. *Journal of Business Venturing*.
- Cooper, A., Gimeno-Gascón, F. J., & Woo, C. Y. (1997). Initial Human and Financial Capital as Predictors of New Venture Performance. *The Journal of Private Equity*.
- Davidsson, P., & Honig, B. (2003). The role of social and human capital among nascent entrepreneurs. *Journal of Business Venturing*.
- Davis, J. P., Eisenhardt, K. M., & Bingham, C. B. (2009). Optimal Structure, Market Dynamism, and the Strategy of Simple Rules. *Administrative Science Quarterly*.
- Espejo, P. G., Romero, C., Hervás, C., & Ventura, S. (2005). Programación genética gramatical para el descubrimiento de reglas de clasificación.
- Ferris, B. (2000). *Nothing Ventured, Nothing Gained*. Allen and Unwin.
- Fried, V. H., & Hisrich, R. D. (1994). Toward a Model of Venture Capital Investment Decision Making. *Financial Management*, *23*(3), 28–37. doi:10.2307/3665619

- Golis, C. (1998). *Enterprise and Venture Capital*.
- Grzech, A. (2009). Preliminary Selection in the Decision-Making Process of Venture Capital Funds, (1).
- Hall, J. (1989). *Venture Capitalist Decision Making and the Entrepreneur*. University of Georgia, Athens.
- Hall, J., & Hofer, C. W. (1993). Venture capitalists' decision criteria in new venture evaluation. *Journal of Business Venturing*.
- He, H. H. H., & Garcia, E. A. (2009). Learning from Imbalanced Data. *IEEE Transactions on Knowledge and Data Engineering*, 21(9). doi:10.1109/TKDE.2008.239
- Hilal, D. K., & Soltan, H. (1992). To prototype or not to prototype? That is the question. *Software Engineering Journal*, 7(6), 388–392.
- Jaffe, A. B., & Trajtenberg, M. (1996). Flows of knowledge from universities and federal laboratories: modeling the flow of patent citations over time and across institutional and geographic boundaries. *Proceedings of the National Academy of Sciences of the United States of America*, 93(23), 12671–12677.
- Kangasharju, A., & Pekkala, S. (2002). The role of education in self–employment success in Finland. *Growth and Change*, 33(Spring), 216–237.
- Leatherbee, M. (2012). Emergence of Entrepreneurial Opportunities: A review on institutions, individuals, and behaviors.
- Magoutas, A. I., Papadogonas, T. A., & Sfakianakis, G. (2012). Market Structure, Education and growth. *International Journal of Business & Social Science*, 3(12), 88–95.
- McMullen, J. S., & Shepherd, D. A. (2006). Entrepreneurial action and the role of uncertainty in the theory of the entrepreneur. *Academy of Management Review*.
- Mitchell, T. M. (1997). *Machine Learning*. *Annual Review Of Computer Science* (Vol. 4, p. 432). McGraw-Hill Science/Engineering/Math. doi:10.1145/242224.242229
- Ries, E. (2011). *The Lean Startup*. Working Paper.
- Riquelme, H., & Rickards, T. (1992). Hybrid conjoint analysis: An estimation probe in new venture decisions. *Journal of Business Venturing*.
- Robledo Castro, E. (2008). *Evaluación de riesgos al otorgar créditos bancarios utilizando árboles de decisión*. Pontificia Universidad Católica de Chile.
- Russell, R. D. (1992). An Examination of the Effects of Organizational Norms, Organizational Structure, and Environmental Uncertainty on Entrepreneurial Strategy. *Journal of Management*.

- Shane, S. (2000). Prior Knowledge and the Discovery of Entrepreneurial Opportunities. *Organization Science*.
- Sierra Araujo, B. (2006). *Aprendizaje automático: conceptos básicos y avanzados* (p. 482). Pearson Prentice Hall.
- Silver, A. D. (1985). *Venture Capital: The Complete Guide for Investors* (New York.). John Wiley and Sons.
- Stalk, G. (1988). Time - The Next Source of Competitive Advantage. *Harvard Business Review*, 66(4), 41–51.
- Sun, Y., Kamel, M. S., Wong, A. K. C., & Wang, Y. (2007). Cost-sensitive boosting for classification of imbalanced data. *Pattern Recognition*, 40(12), 3358–3378.
- Tyebjee, T. T., & Bruno, A. V. (1984). A Model of Venture Capitalist Investment Activity. *Management Science*.
- Van Osnabrugge, M. (1998). Do Serial and Non-Serial Investors Behave Differently?: An Empirical and Theoretical Analysis. *Entrepreneurship: Theory & Practice*, 22(4), 23–42. 20p. 3 Charts.
- Wells, W. A. (1974). *Venture Capital decision making*. Carnegie Mellon University, Pittsburgh.

ANEXOS

ANEXO A: CUESTIONARIO ENVIADO A LOS PARTICIPANTES

Q1 When you applied to Start-Up Chile, how long had your idea been feasible for? (Technically and commercially)

- a. It wasn't feasible on the application date
- b. Less than a year before the application date
- c. One year before the application date
- d. Two years before the application date
- e. Three years before the application date
- f. Four years before the application date
- g. Five or more years before the application date

Q2 What kind of change had caused the emergence of the opportunity when you applied to Start-Up Chile?

- a. A technological change in the inputs that enable your product or solution
- b. A change in the market needs
- c. Both of the above
- d. It is not clear, the opportunity existed a long time ago

Q3 In which industry had you have the most experience when you applied to Start-Up Chile?

- a. Consulting
- b. E-commerce
- c. Education
- d. Energy & Clean Tech
- e. Finance
- f. Healthcare & Biotechnology
- g. IT & Enterprise Software
- h. Import/Export
- i. Mobile & Wireless
- j. Natural Resources - mining, food, lumber, etc.
- k. Media
- l. Social Enterprise
- m. Social Media/Social Network
- n. Tourism
- o. Unclassified establishments

Q4 In which country had you worked the most during the last 3 years, when you applied to Start-Up Chile?

Q5 Which was the first industry that you planned to serve with your startup, when you applied to Start-Up Chile?

- a. Consulting
- b. E-commerce
- c. Education
- d. Energy & Clean Tech
- e. Finance
- f. Healthcare & Biotechnology
- g. IT & Enterprise Software
- h. Import/Export
- i. Mobile & Wireless
- j. Natural Resources - mining, food, lumber, etc,
- k. Media
- l. Social Enterprise
- m. Social Media/Social Network
- n. Tourism
- o. Unclassified establishments

Q6 Which was the first main target market that you planned to launch your startup to, when you applied to Start-Up Chile?

Q7 In what industry did the main strategic resource of your startup resided when you applied to Start-Up Chile?

- a. Consulting
- b. E-commerce
- c. Education
- d. Energy & Clean Tech
- e. Finance
- f. Healthcare & Biotechnology
- g. IT & Enterprise Software
- h. Import/Export
- i. Mobile & Wireless
- j. Natural Resources - mining, food, lumber, etc,
- k. Media
- l. Social Enterprise
- m. Social Media/Social Network
- n. Tourism
- o. Unclassified establishments

Q8 In which country did the main strategic resource of your startup resided when you applied to Start-Up Chile?

Q9 In how many countries did you believed your startup would be operating in 5 following years, when you applied to Start-Up Chile?

- a. Only one country
- b. Two countries
- c. Three countries
- d. Four or more countries

Q10 Which of these statements were you most in agreement with, when you applied to Start-Up Chile?

- a. The main uncertainty is whether the product or service can be developed
- b. The main uncertainty is whether the market is large enough
- c. The startup faces no uncertainties

Q11 What was the highest level of education you had completed when you applied to Start-Up Chile?

- a. Less than High School
- b. High School
- c. 2-Year College Degree
- d. 4-Year College Degree
- e. MBA
- f. Masters (other than MBA)
- g. Ph.D

Q12 How many founders of your startup had finished College studies when you applied to Start-Up Chile?

- a. All
- b. Only two
- c. Only one
- d. None
- e. Don't know

Q13 How many startups had you worked on, when you applied to Start-Up Chile? (either as an employee or the founder)

- a. 0
- b. 1
- c. 2
- d. 3
- e. 4

f. 5+

Q14 How many members of the founder team (including you) had worked on another startup before, when you applied to Start-Up Chile?

- a. 0
- b. 1
- c. 2
- d. 3

ANEXO B: OUTPUT WEKA MODELO 1

```

Test mode: 10-fold cross-validation

=== Classifier model (full training set) ===

CostSensitiveClassifier using reweighted training instances

weka.classifiers.trees.J48 -C 0.25 -M 2

Classifier Model
J48 pruned tree
-----

R3 <= 0: No_Seleccionado (52.15/11.17)
R3 > 0
| R1 <= 0: No_Seleccionado (44.7/8.94)
| R1 > 0
| | R4 <= 0
| | | R2 <= 0: No_Seleccionado (33.52/13.41)
| | | R2 > 0: Seleccionado (32.78/14.9)
| | R4 > 0: Seleccionado (58.85/25.33)

Number of Leaves : 5

Size of the tree : 9

Cost Matrix
0 1
3 0
Time taken to build model: 0.05 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances 149 67.1171 %
Incorrectly Classified Instances 73 32.8829 %
Kappa statistic 0.096
Mean absolute error 0.3897
Root mean squared error 0.428
Relative absolute error 136.2954 %
Root relative squared error 113.6037 %
Coverage of cases (0.95 level) 100 %
Mean rel. region size (0.95 level) 100 %
Total Number of Instances 222

=== Detailed Accuracy By Class ===

          TP Rate  FP Rate  Precision  Recall  F-Measure  MCC  ROC Area  PRC Area  Class
          0,728  0,605  0,854  0,728  0,786  0,102  0,598  0,855  No_Seleccionado
          0,395  0,272  0,231  0,395  0,291  0,102  0,598  0,210  Seleccionado
Weighted Avg.  0,671  0,548  0,747  0,671  0,701  0,102  0,598  0,745

=== Confusion Matrix ===

  a  b  <-- classified as
134 50 | a = No_Seleccionado
 23 15 | b = Seleccionado

```

ANEXO C: OUTPUT WEKA MODELO 2

```

Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

CostSensitiveClassifier using reweighted training instances

weka.classifiers.trees.J48 -C 0.25 -M 2

Classifier Model
J48 pruned tree
-----

R3 <= 0: No_Seleccionado (52.15/11.17)
R3 > 0
|  R1 <= 0: No_Seleccionado (44.7/8.94)
|  R1 > 0
|  |  R4 <= 0: No_Seleccionado (66.3/31.29)
|  |  R4 > 0: Seleccionado (58.85/25.33)

Number of Leaves :    4

Size of the tree :    7

Cost Matrix
  0 1
  3 0

Time taken to build model: 0 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      149           67.1171 %
Incorrectly Classified Instances    73           32.8829 %
Kappa statistic                    0.0495
Mean absolute error                 0.3999
Root mean squared error             0.435
Relative absolute error             139.8476 %
Root relative squared error         115.464 %
Coverage of cases (0.95 level)     100 %
Mean rel. region size (0.95 level) 100 %
Total Number of Instances          222

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                0,745   0,684   0,840     0,745   0,790     0,051   0,574    0,848   No_Seleccionado
                0,316   0,255   0,203     0,316   0,247     0,051   0,574    0,202   Seleccionado
Weighted Avg.   0,671   0,611   0,731     0,671   0,697     0,051   0,574    0,738

=== Confusion Matrix ===

  a  b  <-- classified as
137 47 |  a = No_Seleccionado
 26 12 |  b = Seleccionado

```

ANEXO D: OUTPUT WEKA MODELO 3

```

Test mode:    10-fold cross-validation

=== Classifier model (full training set) ===

CostSensitiveClassifier using reweighted training instances

weka.classifiers.trees.J48 -C 0.25 -M 2

Classifier Model
J48 pruned tree
-----

R3 <= 0: No_Seleccionado (52.15/11.17)
R3 > 0
 |  R1 <= 0: No_Seleccionado (44.7/8.94)
 |  R1 > 0
 | |  R2 <= 0: No_Seleccionado (74.5/35.76)
 | |  R2 > 0: Seleccionado (50.66/21.6)

Number of Leaves   :    4

Size of the tree   :    7

Cost Matrix
  0 1
  3 0
Time taken to build model: 0.02 seconds

=== Stratified cross-validation ===
=== Summary ===

Correctly Classified Instances      148           66.6667 %
Incorrectly Classified Instances     74           33.3333 %
Kappa statistic                     0.0447
Mean absolute error                  0.404
Root mean squared error              0.4366
Relative absolute error              141.2716 %
Root relative squared error          115.8848 %
Coverage of cases (0.95 level)      100          %
Mean rel. region size (0.95 level)  100          %
Total Number of Instances           222

=== Detailed Accuracy By Class ===

                TP Rate  FP Rate  Precision  Recall  F-Measure  MCC      ROC Area  PRC Area  Class
                0,739   0,684   0,840     0,739   0,786     0,047   0,555   0,843   No_Seleccionado
                0,316   0,261   0,200     0,316   0,245     0,047   0,555   0,190   Seleccionado
Weighted Avg.   0,667   0,612   0,730     0,667   0,693     0,047   0,555   0,731

=== Confusion Matrix ===

  a  b  <-- classified as
136 48 |  a = No_Seleccionado
 26 12 |  b = Seleccionado

```