



UNIVERSITAT OBERTA DE CATALUNYA (UOC)
MÁSTER UNIVERSITARIO EN CIENCIA DE DATOS (*Data Science*)

TRABAJO FINAL DE MÁSTER

ÁREA: 1

Aplicación de aprendizaje automático para la predicción de empresas de alto crecimiento en Colombia

Autor: Cristhyan Leonardo Naranjo Puertas

Tutor: Albert Solé Ribalta

Profesor: Santiago Rojo Muñoz

Bogotá, 16 de enero de 2022



Esta obra está sujeta a una licencia de Reconocimiento - NoComercial - SinObraDerivada
3.0 España de Creative Commons.

Copyright ©2022 Cristhyan L. Naranjo Puertas

FICHA DEL TRABAJO FINAL

Título del trabajo:	Aplicación de aprendizaje automático para la predicción de empresas de alto crecimiento en Colombia
Nombre del autor:	Cristhyan Naranjo Puertas
Nombre del colaborador/a docente:	Albert Solé Ribalta
Nombre del PRA:	Santiago Rojo Muñoz
Fecha de entrega :	12/2021
Titulación o programa:	Máster Universitario en Ciencias de Datos
Área del Trabajo Final:	Trabajo Final De Máster
Idioma del trabajo:	Español
Palabras clave	Ciencia de datos, Empresas de alto crecimiento, Aprendizaje supervisado

Dedicatoria/Cita

A mi hijo , espero que este trabajo te sirva de inspiración, y a mi familia por todo su apoyo

Resumen

Las Empresas de alto crecimiento (EAC) son responsables de aproximadamente el 50 % del total de ventas y la generación de nuevos empleos en la economía de un país, por esta razón siempre ha sido uno de los principales intereses dentro de la demografía empresarial poder distinguirlas y entender su ciclo de vida. Hasta hace poco los resultados usando los métodos clásicos no parecían alentadores, no fue hasta el auge de los modelos de aprendizaje automatizado con su alta capacidad de encontrar relaciones no evidentes entre las variables, que se está logrando generar modelos que logran predecir si una empresa llegara a ser de alto crecimiento con un nivel alto de exactitud. En el caso de Colombia no se han realizados estudios profundos en la identificación y predicción de EAC dentro de su economía particular, en vista de esto, con este trabajo se buscó poder encontrar los predictores con mayor relevancia dentro de la economía colombiana al momento de pronosticar si una empresa lograra convertirse en EAC y usarlos en la elaboración de un modelo de aprendizaje automatizado que logre predecir con exactitud si una empresa tiene el potencial de llegar a ser de alto crecimiento. Luego de generar varios modelos a partir de ensayos con diferentes variables se encontró que efectivamente es posible conseguir altos niveles de exactitud entrenando un modelo de random forest a partir de información obtenida de los balances financieros de las empresas y que en general los predictores identificados coinciden con los encontrados relevantes en investigaciones previas.

Abstract

High Growth Companies (HGC) are responsible for approximately 50 % of total sales and the generation of new jobs in the economy of a country, for this reason it has always been one of the main interests within business demography to be able to distinguish them and understand its life cycle. Until recently the results using classical methods did not seem encouraging, it was not until the rise of machine learning models with their high capacity to find non-obvious relationships between variables, that it is possible to generate models that manage to predict if a company to be high growing with a high level of accuracy. In the case of Colombia, no in-depth studies have been carried out on the identification and prediction of HGC within its particular economy, in view of this, with this work it was sought to be able to find the most relevant predictors within the Colombian economy at the time of forecasting. if a company were able to become HGCs and use them to build a machine learning model that accurately predicts whether a company has the potential to become high-growth. After generating several models from tests with different variables, it was found that it is indeed possible to achieve high levels of accuracy by training a random forest model from information obtained from the financial balances of the companies and that, in general, the identified predictors coincide with those found relevant in previous research.the predictors found coincide with those found to be relevant in previous research.

Palabras clave: Ciencia de datos, Empresas de alto crecimiento, Aprendizaje supervisado

Índice general

Abstract	VII
Índice	IX
Listado de Figuras	XI
Listado de Tablas	1
1. Introducción	3
2. Estado del arte	5
2.0.1. Definiciones de empresa de alto crecimiento	5
2.0.2. ¿Qué trabajos previos existen?	7
3. Procesamiento previo	9
3.0.1. Datos de las empresas y balances financieros	9
3.0.2. Generación de las etiquetas de EAC	14
4. Generación de modelos	17
4.0.1. Ensayos con otros modelos	22
4.0.2. Importancia de predictores	23
5. Conclusiones	27
Bibliografía	28
6. Anexos	31
6.1. Variables de los balances financieros	31

Índice de figuras

2.1. Diagrama del calculo de EAC según CONFECAMARAS. Elaboración propia con base en la imagen dentro de la publicación	7
3.1. Frecuencia de empresas por sector económico	12
3.2. Frecuencias de nacimiento de empresas por año.	12
3.3. Mapa coropletico de la cantidad de empresas por departamento	13
3.4. Matriz de correlación a partir del método de Pearson	15
3.5. Relación entre las variables solvencia y endeudamiento	16
3.6. Matriz de correlación a partir del método de Spearman	16
4.1. Esquema del funcionamiento del método smote para la generación de registros sintéticos	18
4.2. Árbol de selección generado a partir de la base nivelada	19
4.3. Resultado de la búsqueda del óptimo de penalización usando cross-validation	22
4.4. Curva ROC para el modelo de red neuronal	23
4.5. Importancia de los predictores usando el método de la pureza de nodos	24
4.6. Importancia de los predictores usando el método de permutación	24

Índice de cuadros

3.1. <i>Bases usadas durante el ejercicio.</i>	10
3.2. <i>Porcentaje de empresas de alto crecimiento según la metodología utilizada</i>	15
4.1. <i>Resultados de los modelos con las nuevas variables.</i>	21

Capítulo 1

Introducción

Dentro de las actividades encaminadas en el desarrollo de la economía de un país se encuentra la constante búsqueda de nuevas fuentes de generación de empleos y de riqueza que contribuyan al avance económico de la nación, también en esa misma dirección se encuentran inversionistas privados en búsqueda de empresas que en un corto tiempo tengan un alto crecimiento de tal forma que además de garantizar el éxito de sus inversión, puedan obtener ganancias de forma acelerada para poder realizar con ellas nuevas inversiones. Estas entre otras razones han impulsado el estudio de las actualmente denominadas Empresas de Alto Crecimiento (EAC) (y de su subconjunto de empresas gacela) con el fin de poder identificar tempranamente que compañías tienen el potencial de convertirse en una de estas empresas y así a partir de ayudas del gobierno o inversiones privadas sean impulsadas a lograr esa meta.

Ahora bien cualquier persona que halla iniciado un proyecto sabe que por más planificación y por más que se trate de controlar todas las variables que puedan afectar el desarrollo del proyecto como se tenía presupuestado, siempre existen situaciones con comportamiento aleatorio que pueden comprometer los planes, mas aun, dependiendo del nivel de perturbación y en que parte del desarrollo del proyecto se origine, bajo las condiciones de un efecto mariposa (sistemas caóticos) no es necesario una gran perturbación para que cambie por completo el desarrollo del proyecto, por ejemplo, el cambio de una legislación, cosa que no esta bajo el control de los emprendedores, podría hacer que una empresa al no poder adaptarse a las nuevas políticas se vaya a la quiebra, del mismo modo que podría generarle ventajas que la impulsen en un alto crecimiento. Teniendo esto en cuenta es claro ver que el desarrollo de la vida de una empresa esta influenciado además de, por las variables que son controlables, también son afectadas por situaciones fuera de su control que hace que no sea fácil poder generar modelos que logren predecir el nivel de crecimiento de una organización a través del tiempo.

En la actualidad , dado el aumento de nuevas fuentes de información, por ejemplo el uso de registros administrativos con cada vez más calidad y oportunidad en su información, y el

auge del uso del aprendizaje automatizado, ha permitido que aprovechando la capacidad que tienen estos modelos de encontrar relaciones entre las variables que identifican el estado de una empresa, las cuales bajo el marco de estudios de econometría clásicos no eran relevantes o no eran fácilmente identificables, se logre caracterizar estados que propician que una empresa llegue a tener un alto crecimiento.

Teniendo en cuenta que las economías entre países son muy variadas y diferentes, razón por lo que aun se encuentra en desarrollo una definición general de lo que es una empresa de alto crecimiento y el modo en que se realizan los cálculos de este indicador, por el momento se ha visto que, aun sacrificando la capacidad de comparar los resultados entre naciones, es mejor tener parámetros de clasificación adaptados según la economía de cada país. Con este trabajo se buscó además de contribuir con el poco estudio que existen de EAC dentro de la economía local, generar un modelo de clasificación que en el marco de la economía colombiana fuera capaz de predecir a partir de la información de balances financieros del año 2017 si una empresa lograría en un periodo de 3 años en el futuro convertirse en empresa de alto crecimiento. Con este modelo se espera que en un futuro se logre utilizando información reciente, identificar pequeñas y medianas empresas que se encuentren en un estado temprano en el proceso de convertirse en EAC y con esto buscar un patrocinio ya sea privado o público que aliente ese crecimiento, para así de esta forma garantizar en un futuro cercano la generación de nuevos empleos y un aumento en la productividad que ayude en la evolución económica de la nación.

Capítulo 2

Estado del arte

2.0.1. Definiciones de empresa de alto crecimiento

A partir de los estudios realizados por David Birch a finales de los años 70 se mostró que, a diferencia del pensamiento generalizado, la mayoría de nuevos empleos son creados por pequeñas y medianas empresas, en su trabajo *The job generation process*[\[4\]](#) demuestra como en Estados Unidos 2 tercios de los nuevos empleos entre 1969 y 1976 fueron creados por empresas de menos de 20 empleados, en la actualidad estudios han mostrado que esta tendencia se ha mantenido, por ejemplo el Banco Mundial en un estudio realizado en el 2019[\[6\]](#) mostró que aunque las EAC conforman apenas entre el 3 % al 20 % de las empresas de manufacturas y servicios, estas generan más del 50 % de todos los nuevos empleos y de las ventas en estos sectores combinados. En vista de la clara importancia de estas empresas dentro de la economía de un país, los economistas que trabajan en el área de demografía empresarial han buscado como identificar las principales características que hacen que una empresa logre ese nivel de crecimiento, en este sentido se ha venido madurando el concepto de empresas de alto crecimiento (EAC) , que a groso modo son empresas que han aumentado de manera significativa y continua su número de empleados o de ventas en un periodo corto de tiempo. Con el fin de poder comparar resultados y de tener una definición general se han propuesto varias condiciones que permitan clasificar a una empresa como EAC, actualmente una de las definiciones mas usadas es la propuesta por la OCDE[\[1\]](#):

Todas las empresas con un crecimiento medio anualizado superior al 20 % durante un periodo de tres años deben considerarse como empresas de alto crecimiento. El crecimiento puede medirse por el número de empleados o por el volumen de negocio.

Bajo esta definición, para identificar una EAC basta con considerar únicamente la población de empresas activas que alcanzan el umbral definido de empleados (se sugiere que sea de por

lo menos 10 empleados) en el año $t-3$, medir el número de empleados en el año t y aplicar la siguiente ecuación:

$$\sqrt[3]{\frac{\text{empleados}_{(t)}}{\text{empleados}_{(t-3)}}} - 1 > 0,2$$

Medido en volumen de ventas

$$\sqrt[3]{\frac{\text{ventas}_{(t)}}{\text{ventas}_{(t-3)}}} - 1 > 0,2$$

Con base en esta definición, el Comité de Expertos en Estadísticas de Negocios y Comercio de las Naciones Unidas (UNCEBTS, por sus siglas en inglés) dentro de sus *Fichas metodológicas para el conjunto principal de indicadores de negocios y comercio*[9], decide optar por un porcentaje de crecimiento menos restrictivo del 10 % (esta misma definición es usada por la UNECE en su *Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics*[2])

$$\sqrt[3]{\frac{\text{empleados}_{(t)}}{\text{empleados}_{(t-3)}}} - 1 > 0,1$$

Medido en volumen de ventas

$$\sqrt[3]{\frac{\text{ventas}_{(t)}}{\text{ventas}_{(t-3)}}} - 1 > 0,1$$

La razón principal de que no se tenga una definición general de EAC es el hecho de que la economía de cada país es diferente y no todos los países tienen acceso a la información necesaria para el cálculo de este indicador. Un ejemplo de definición local de EAC la ofrece *Confecámaras* en su *Manual de demografía empresarial adaptado para Colombia*[7], donde propone con base en la economía local del país que se clasifique como EAC a las empresas que al inicio del periodo tienen por lo menos 10 empleados y mantienen una variación porcentual positiva entre años superior al 20 % en el número de empleados o de ventas.



Figura 2.1: Diagrama del calculo de EAC según CONFECAMARAS. Elaboración propia con base en la imagen dentro de la publicación

2.0.2. ¿Qué trabajos previos existen?

Dada su importante participación en el crecimiento económico y la generación de nuevos empleos, Las EAC han sido altamente estudiadas con el fin de encontrar características que permitan diferenciarlas o mejor aún poder identificar condiciones iniciales con las que se pueda lograr predecir si una empresa tiene el potencial de llegar a ser una EAC en un futuro. Hasta hace poco los avances en esta área no fueron muy prometedores, por un lado, dado que la economía de cada país tiene características propias, como su nivel de industrialización, distribución geográfica, nivel de empleo formalizado entre otros, hace que sea muy complicado identificar variables generales que permitan predecir el comportamiento en el crecimiento de cualquier empresa en el mundo[3], además los modelos econométricos utilizados han mostrado que la evolución de una empresa depende de, además de factores invariantes en el tiempo (o de lenta variación) como la orientación de la empresa o el nivel educativo de los propietarios, también depende de factores con comportamiento aleatorio que hacen que la predicciones no logren ser precisas, este comportamiento errático llega al punto de que se puede lograr una buena aproximación usando modelos basados en la *Ruina del apostador* que es un caso especial de una caminata aleatoria.[5]

No fue solo hasta la llegada de la revolución del aprendizaje automatizado que ha sido posible entrenar modelos capaces de, a partir de grandes cantidades de datos, encontrar relaciones entre las variables que no son evidentes con los métodos tradicionales. De esta forma se ha logra-

do identificar variables relevantes en la formación de EAC y generar modelos con niveles de exactitud altos, por ejemplo dentro de la compilación *Supervised Learning for the Prediction of Firm Dynamics*[3] que reúne experiencias con modelos de aprendizaje automatizado aplicados al estudio de dinámicas y demografía empresarial, se encontró que en los casos relacionados con EAC, en los diferentes escenarios de estudio se logró bondades de ajuste en los modelos suficientemente buenos para ser considerados como viables en la predicción de este tipo de empresas, más aún en el estudio *Is firm growth random? A machine learning perspective* [10] se evaluó cual modelo lograba el mejor ajuste dentro de varios modelos diferentes entrenados usando los mismos datos, encontrando que con el modelo de bosques aleatorios (random forest) se conseguía el mejor rendimiento.

En Colombia no se han realizado muchos estudios sobre las EAC nacionales ya que en general los datos empresariales necesarios no cuentan con la calidad necesaria, por ejemplo se necesita controlar la continuidad de las empresas durante los periodos de estudio, para esto se debe tener información de la supervivencia de la empresa lo que para algunas no es posible por situaciones como, por ejemplo, que aunque la empresa no se encuentre activa sigue apareciendo en los registros, o también que habiendo sucedido un cambio de locación geográfica o de razón social, situaciones que se deben de interpretar como una muerte y nacimiento empresarial lo que causa que se pierda la continuidad de la empresa, su número de identificación tributario sigue siendo el mismo aparentando una persistencia en su supervivencia, estas entre otras razones impiden que se pueda tener una base de datos confiable. Aun así existen investigaciones recientes en donde se realiza un estudio paralelo entre empresas Españolas y Colombianas[8] buscando caracterizar las variables principales que definen a una EAC en cada país, este estudio aunque fue realizado con métodos tradicionales mostró que efectivamente las condiciones regionales hacen que sea necesario generar modelos particulares para cada situación y que ciertamente el tamaño inicial de la empresa, su edad, su locación o la rama de actividad económica a la que pertenecen son variables que influyen en la posibilidad de que una empresa llegue a ser una EAC.

Capítulo 3

Procesamiento previo

La fuente principal para la elaboración del ejercicio proviene de los balances generales de las empresas presentes en los registros de la superintendencia de sociedades entre los años 2017 a 2020, los cuales al ser datos de dominio público y ser información recopilada por un ente de regulación nacional, se garantiza la veracidad de la información y se conserva un esquema de los datos continuo durante el periodo de estudio lo que facilita su tratamiento, también se realizaron ensayos añadiendo variables provenientes de bases abiertas con información de contratos con el estado, modificaciones en los estatutos de las entidades e incidentes judiciales que en principio pueden mejorar la calidad del modelo al contener información relevante para la clasificación de EAC.

Tanto los procesos previos de exploración y adecuación de los datos, como el entrenamiento y evaluación de los diferentes modelos se realizó bajo el lenguaje R.

Para la elaboración de este ejercicio, como primer paso se cargaron los diferentes archivos en el sistema para constatar que ninguna de las bases estuviera corrupta o presentara alguna anomalía (por ejemplo, columnas corridas o formatos no reconocibles), los resultados de la carga se encuentran en la tabla 1

3.0.1. Datos de las empresas y balances financieros

Se usó la base *daem* como tabla maestra para, a partir de ella construir las diferentes combinaciones de variables que serán usadas en la construcción de los diferentes modelos. Para esto se revisó la existencia de registros duplicados a partir de la unicidad del indicador sintético *ID* para el que se encontraron 13 registros duplicados de los cuales 11 fueron eliminados. Para la revisión de las demás variables, primero se comprobó su nivel de completitud encontrando que hay o no datos vacíos en el identificador, la variable CIU (Clasificación Industrial Internacional Uniforme) y la de departamento (localización geográfica) que en la literatura son

Nombre del archivo	Nombre en la base	Descripción	Cantidad de registros	Cantidad de columnas
Datos-Empresa.xlsx	daem	Datos basicos de las empresas	12279	17
Bal2017.xlsx	bal17	datos de balances del 2017	12279	48
Bal2018.xlsx	bal18	datos de balances del 2018	12279	48
Bal2019.xlsx	bal19	datos de balances del 2019	12279	48
Bal2020.xlsx	bal20	datos de balances del 2020	11891	49
ModificacionEstatutos.xlsx	moes	Número modificaciones de los estatutos de la empresa registrados (modificaciones de capital, nombramientos/revocaciones, cambios de forma jurídica, modificación de objeto social, etc)	29374	3
RamaJudicial.xlsx	raju	Número de incidencias judiciales publicadas por los juzgados y tribunales de Colombia	1881	3
ContratosEstado.xlsx	coes	Número de contratos públicos (licitaciones) ganados/obtenidos y su valor aproximado	85924	4

Cuadro 3.1: Bases usadas durante el ejercicio.

identificadas como variables importantes, mientras que las variables empleados y constitución que serán usadas para determinar que empresas cumplen con las condiciones de la definición de EAC, tienen 286 y 16 valores faltantes respectivamente, al ser una cantidad de registros relativamente pequeña se decidió retirarlos.

Por la definición de EAC se tiene que no se puede realizar el ejercicio con empresas de menos de 5 años, esto es nacidas entre t y $t+4$ del periodo de estudio, por esta razón se retiran las empresas que estén dentro de ese intervalo (empresas muy Jóvenes), además se mantuvo solo las empresas que tuvieran en la variable estado *Activa* para evitar la perdida de continuidad de alguna empresa. Al final se pasó de 12279 registros iniciales a 11862 correspondiendo a una pérdida del 3.4 % de la información original.

Continuando con las variables importantes se encontró que en la variable Empleo, 1386 empresas (un 8.55 % del total) son considerados outliers, pero dado que dentro de la diversidad de las empresas presentes en la base se encuentran organizaciones grandes de varios años de actividad junto con empresas jóvenes que no tienen una gran cantidad de empleados, es de esperarse la presencia de estos valores extremos. Al realizar la gráfica de caja sin representar los valores extremos se ve como el grueso de la variable está por debajo de 124 empleados y que existen empresas con una cantidad muy alta de empleados comparados con el resto de las empresas dentro de esta base. Según la literatura, esta variable es importante dentro del estudio porque se ha visto que empresas con muchos empleados están bien establecidas dentro del mercado y tienen un cubrimiento grande del país, pero estas empresas también pueden tener una tasa de crecimiento menor comparada con una empresa joven en donde un aumento pequeño en la cantidad de empleados es significativo. Para poder aprovechar esta variable se discretizo dejando 10 categorías a partir de sus deciles.

Para la variable CIIU, la OCDE recomienda la exclusión de empresas pertenecientes a las secciones A,O,T y U, como Colombia tiene una industria importante de agricultura, no se debe de retirar la sección A. Al explorar las secciones CIIU se encontró que la mayoría de empresas se encuentran en las secciones G: *Comercio al por mayor y al por menor; reparación de vehículos automotores y motocicletas*, C: *Industria manufacturera* y L: *Actividades inmobiliarias*. Además se excluyeron las secciones: K *Actividades financieras y de seguros* U: *Actividades de organizaciones y entidades extraterritoriales o administración pública y defensa; planes de seguridad social de afiliación obligatoria*, también se excluyen las empresas sin ánimo de lucro, que mayoritariamente se encuentran en el CIIU 9499: *Actividades de otras asociaciones n.c.p.*, al revisar estos casos por su tipología y objeto social solamente el registro 7550 se debe de

excluir. Por ultimo Se pensó en eliminar los holdings empresariales (correspondientes al CIIU 6491: *Leasing financiero*) pero no se encontró ninguna.

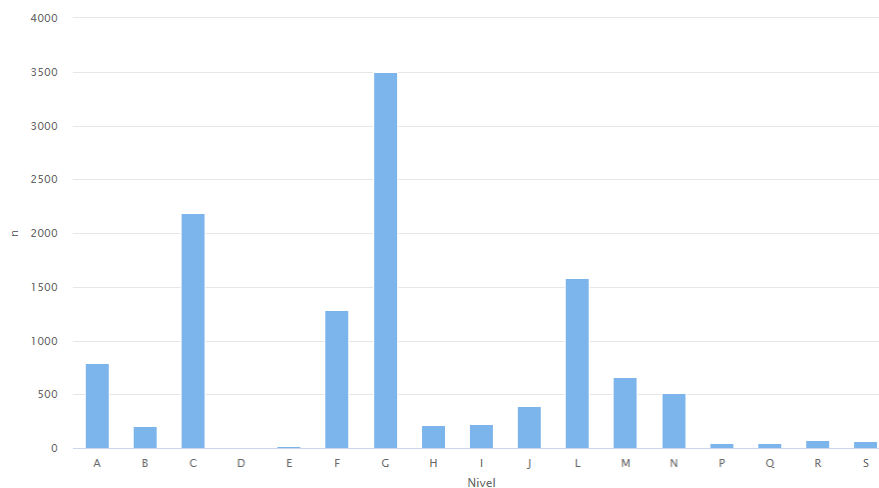


Figura 3.1: Frecuencia de empresas por sector económico

Al revisar la variable constitución correspondiente al año de nacimiento de la empresa se encontró que hubo un crecimiento muy constante en el nacimiento de empresas hasta llegar a un máximo en el 2011, desde donde comienza una caída en la creación de nuevas empresas hasta el año 2015. Para utilizar esta variable dentro del entrenamiento se discretizo a partir de sus deciles.

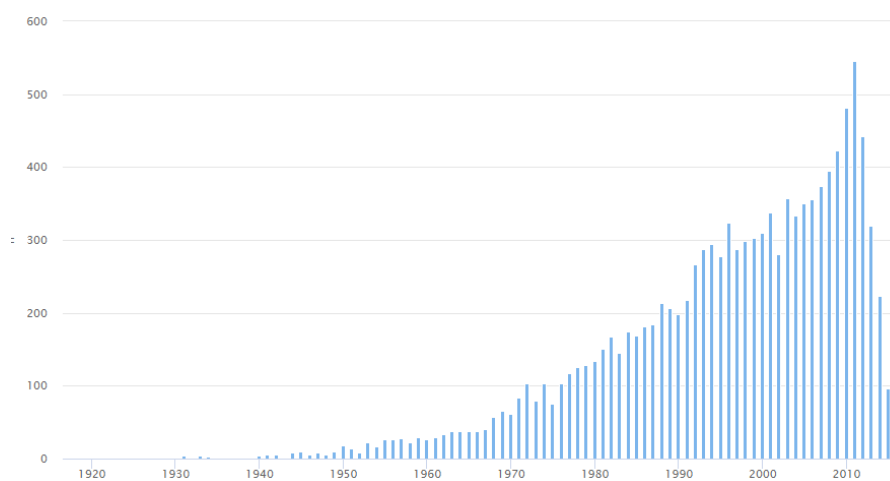


Figura 3.2: Frecuencias de nacimiento de empresas por año.

Otra de las variables que la literatura describe como significativa para que una empresa

llegue a convertirse en EAC es su ubicación geográfica, para buscar un entrenamiento optimo se tomo la localización geográfica a nivel de departamento, además para visualizar mejor la información se genero un mapa coropletico con la división de colores determinados a partir de los quintiles de la cantidad de empresas por departamento

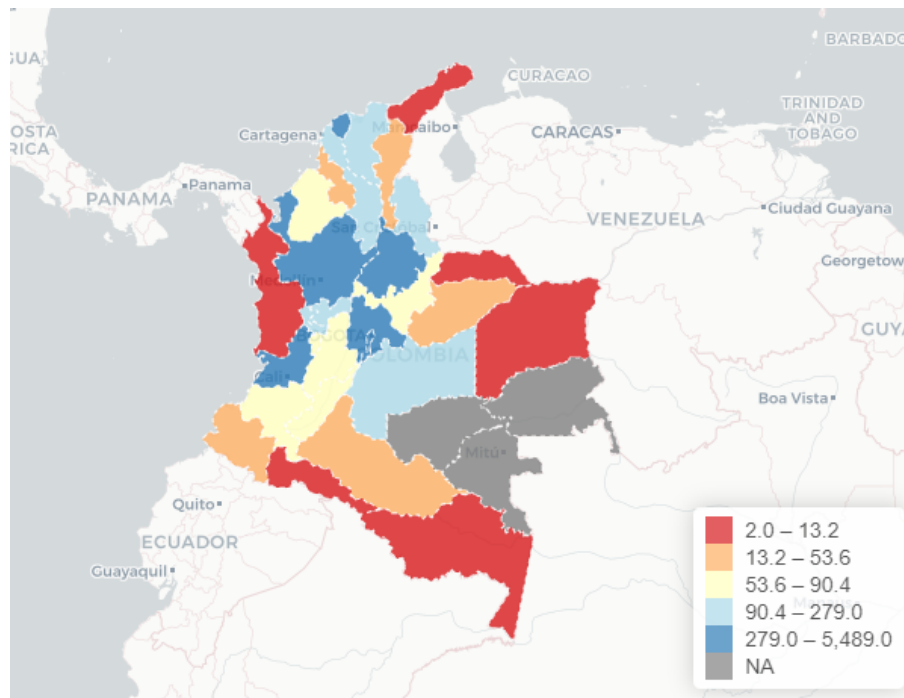


Figura 3.3: Mapa coropletico de la cantidad de empresas por departamento

Claramente en el mapa se ve como la región oriental, la Amazonia, el Choco y la Guajira tienen una cantidad muy pequeña de empresas, esto es de esperarse ya que estas regiones presentan baja cantidad de población, además de un continuo abandono por parte del estado, por otro lado regiones como Bogotá, Cundinamarca, Antioquia, Atlántico y el Valle, al contener las capitales principales del país muestran una alta cantidad de empresas, esto respalda la bibliografía que afirma que las empresas de alto crecimiento (y empresas en general) se ubican en ciudades principales.

Los datos que agregan la dimensión longitudinal en el estudio se encuentran dentro de las bases de balances, estos datos son los que permiten tener una evolución en el tiempo de las empresas, En una primer revisión se detecto que dentro del balance del año 2020 se encontraba una columna extra llamada BASICO que no estaba presente en los demás años por lo que se decidió eliminarla, también al revisar la continuidad de las empresas dentro de las bases se vio que en los años 2017 a 2019 los balances contienen las mismas empresas, pero en el año 2020 desaparecieron 387, probablemente causado por la caída económica generada por la pandemia,

para mantener la continuidad de los datos se decidió eliminar estas empresas de los años anteriores. Vale aclarar que a pesar de que el año 2020 por sus condiciones particulares, en donde la economía no pudo tener un comportamiento normal y las empresas fueron fuertemente afectadas disminuyendo su productividad por la ausencia de trabajadores y de consumo, se decidió mantener la información de ese año con el fin de por un lado cumplir con las condiciones necesarias que se piden en la definición de EAC y de además, buscar un modelo con un nivel de robustez que le permita trabajar en condiciones más generales.

Las bases de balances contienen una cantidad significativa de variables las cuales fueron utilizadas para obtener nuevas variables más generales y relevantes dentro del ambiente financiero (esto además permite que las variables más generales condensen la información de las variables con la que se elaboran, realizando así una manera de reducción de dimensionalidad), las relaciones entre las variables que se usaron en las construcción de las nuevas variables: resultado del ejercicio, tesorería, liquidez, solvencia, endeudamiento, rentabilidad, fondo de maniobra y ratio de cobertura de interés se puede ver en el anexo 1

Con los datos de la base *daem* y los de los balances, ya es posible realizar la etiqueta que define si una empresa es EAC o no, la cual será usada durante el entrenamiento supervisado. Para este ejercicio, basado en los datos disponibles se decidió definir las EAC a partir del crecimiento en los Ingresos operacionales (variable RAIE), aunque también en paralelo se realizaron pruebas con las variables resultado del ejercicio y resultados antes de impuestos. Aprovechando las relaciones entre estas variables de la base de balances se buscó imputar los valores faltantes para los casos en que fuera posible construirlas a partir de otras, en los casos en que no fue posible y al ver que se contaban con una cantidad de registros suficientemente grande se decidió dejar solo las empresas que contaban con toda la información para así tener una base completa con la cual realizar los entrenamientos.

3.0.2. Generación de las etiquetas de EAC

Como se pudo ver en la revisión de literatura, si bien la definición de EAC dada por la OCDE se ha popularizado dentro de los institutos de estadísticas para la realización del cálculo de este indicador, aun no existe una definición definitiva, además las pruebas han mostrado que dadas las particularidades de la economía de cada país no es adecuado generalizar una definición, con base en esto se decidió realizar una prueba con diferentes metodologías para ver su comportamiento sobre los datos de las empresas que se están estudiando. Se probaron 2 diferentes metodologías, la primera evalúa un crecimiento continuo superior al 10 % y 20 % en la variación porcentual entre los años 2017 a 2020, este método tiene la ventaja de tener en cuenta los valores de todos los años. La siguiente es de un crecimiento medio anualizado supe-

rior al 10 % y 20 %, que es la metodología sugerida por la OCDE, la cual tiene la desventaja de solo operar con los valores del primer y último año (que en este caso es el 2020, un año con valores atípico por los efectos de la pandemia). Como resultado se obtuvieron las siguientes proporciones de EAC presentes en la base:

Metodología	Porcentaje
Variación porcentual con el 10 %	4.49 %
Variación porcentual con el 20 %	1.40 %
Crecimiento medio anualizado con el 10 %	21.28 %
Crecimiento medio anualizado con el 20 %	9.28 %

Cuadro 3.2: *Porcentaje de empresas de alto crecimiento según la metodología utilizada .*

Si bien con cada método se consiguieron proporciones razonables, se optó por continuar el proceso usando el de la variación porcentual sobre un umbral del 10 %.

Después de haber realizado la etiquetación de las empresas, se continuo con la revisión de las variables que fueron construidas, primero las variables rentabilidad y ratio de cobertura de intereses están relacionadas con las variables usadas para etiquetar las empresas por lo que no se usaran como predictores. Luego se realizó un análisis de correlación (bajo el método de *Pearson*) de las variables por medio del factor de inflación de la varianza (que mide el nivel de multicolinealidad).

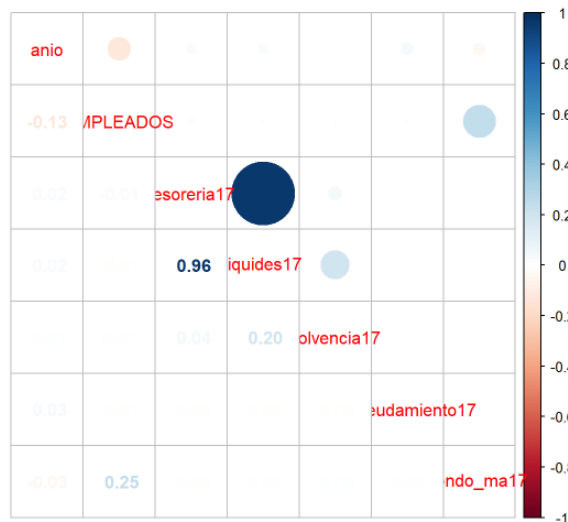


Figura 3.4: Matriz de correlación a partir del método de Pearson

Se encontró que tanto el factor de inflación de varianza, como la matriz de correlaciones muestran que existe una relación entre las variables de tesorería y liquidez, donde por su nivel

de completitud dentro de la base, se decidió solo se dejar la última de estas variables.

Al realizar una inspección de las relaciones entre las variables a partir de sus gráficas se encontró que la solvencia y el endeudamiento mostraban una relación no lineal decreciente por lo que se calculó una correlación de las variables usando esta vez el método de *Spearman*, donde se comprobó que efectivamente existe la correlación, Viendo esto se decidió dejar solo la variable de endeudamiento.

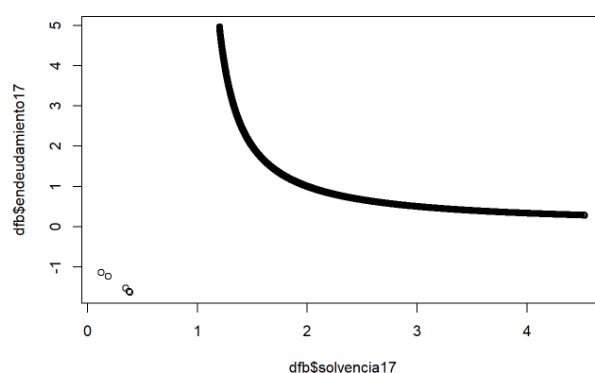


Figura 3.5: Relación entre las variables solvencia y endeudamiento

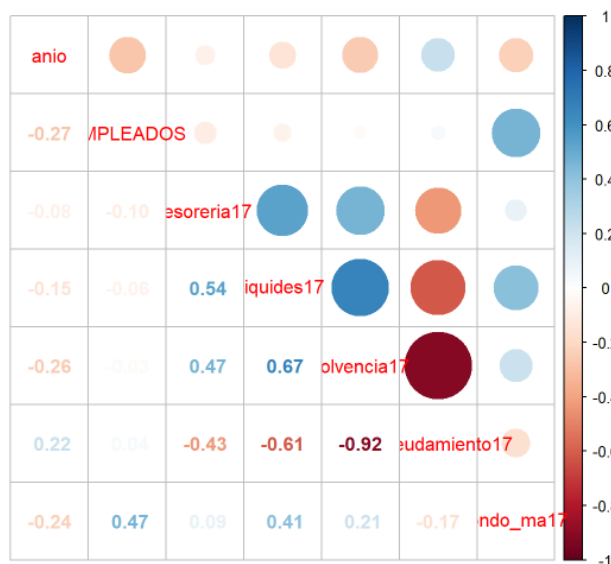


Figura 3.6: Matriz de correlación a partir del método de Spearman

Capítulo 4

Generación de modelos

Teniendo etiquetadas las empresas que al cabo del periodo entre 2017 a 2020 lograron llegar a ser EAC, se propone encontrar un modelo de clasificación que a partir de solo los datos del año 2017 sea capaz de predecir si una empresa se convertirá en EAC dentro del periodo de 3 años en el futuro.

El primer ensayo para la generación de un modelo de predicción se realizó a partir del entrenamiento de un único árbol aleatorio, ya que este cuenta con la posibilidad de visualizar sus ramas y nodos lo que permite darse una primera idea de que variables influyen más durante el proceso de clasificación, para esto fue necesario que los datos categóricos se transformaran a variables tipo factor (con esto, salvo en el entrenamiento de redes neuronales, se le indica al sistema que se trata de variables categóricas y no es necesario aplicar un one hot encoding), además se eliminó los outliers de las variables de liquides, endeudamiento y fondo de maniobra para continuar con una normalización de los datos numéricos.

Al generar el primer árbol con una profundidad mínima de 10 niveles y una base de entrenamiento del 70% del total de los datos, se encontraron 2 dificultades en el proceso de entrenamiento y evaluación del modelo, la primera es que dado que la cantidad de EAC presentes en la base es significativamente pequeña (como se vio es del 4.5%) los datos presentan un alto desbalance que hace que el modelo aprenda muy bien a reconocer la clase mayoritaria compuesta por las empresas que no son EAC, porque no tiene suficientes insumos para aprender a clasificar las que si son EAC, la segunda cuestión es que al evaluar el modelo con los indicadores tradicionales para modelos de clasificación se encuentra que, por ejemplo el F1 score arroja una precisión del 97%, lo cual para nuestro ejercicio es engañoso dado que nuestro interés no es clasificar las empresas que no son EAC. Así para solucionar el desbalance de los datos se aplicó el método de *smote* (en este caso particular se usa *smote_nc*[\[11\]](#) que puede trabajar también con variables categóricas, utilizando par este tipo de variables la distancia de Gower), que genera variables sintéticas de la categoría más pequeña al tomar elementos que

se encuentren entre un punto de la categoría y alguno de sus vecinos cercanos (calculados con KNN).

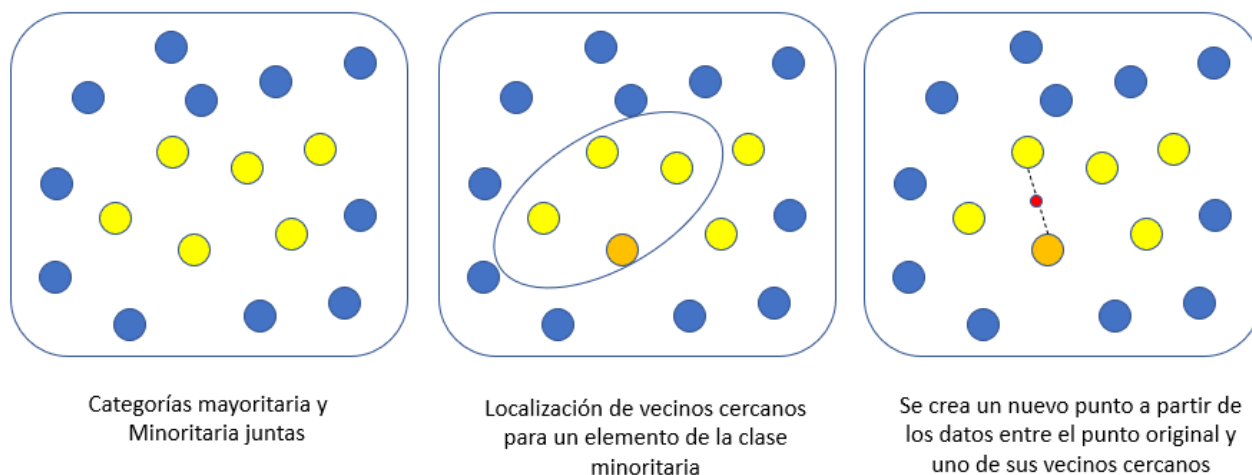


Figura 4.1: Esquema del funcionamiento del método smote para la generación de registros sintéticos

Al aplicar el método de *smote* con un aumento de la clase minoritaria del 70 % de respecto de la clase mayoritaria y un k de 5 para el KNN, se pasa a tener una base con 6904 empresas no EAC y 367 empresas EAC, a tener 6904 empresas no EAC y 4833 empresas EAC (también es posible nivelar la base al realizar una submuestra retirando registros de la clase mayoritaria). Para el problema de la evaluación del modelo se optó por usar como indicador el de *especificidad* ya que este solo calcula el ratio de aciertos en la predicción de la categoría de interés.

Al repetir el entrenamiento del árbol con la base balanceada usando la misma configuración se obtuvo una especificidad del 0.8062, donde el principal discriminante fue el de pertenecer a algunas secciones económicas (entre las cuales es pertenecer a la clase G o C) luego el de ser una empresa con una cantidad baja de empleados (con un máximo 15 empleados), y edad favoreciendo a las empresas jóvenes, criterios que van acorde a lo esperado según la bibliografía.

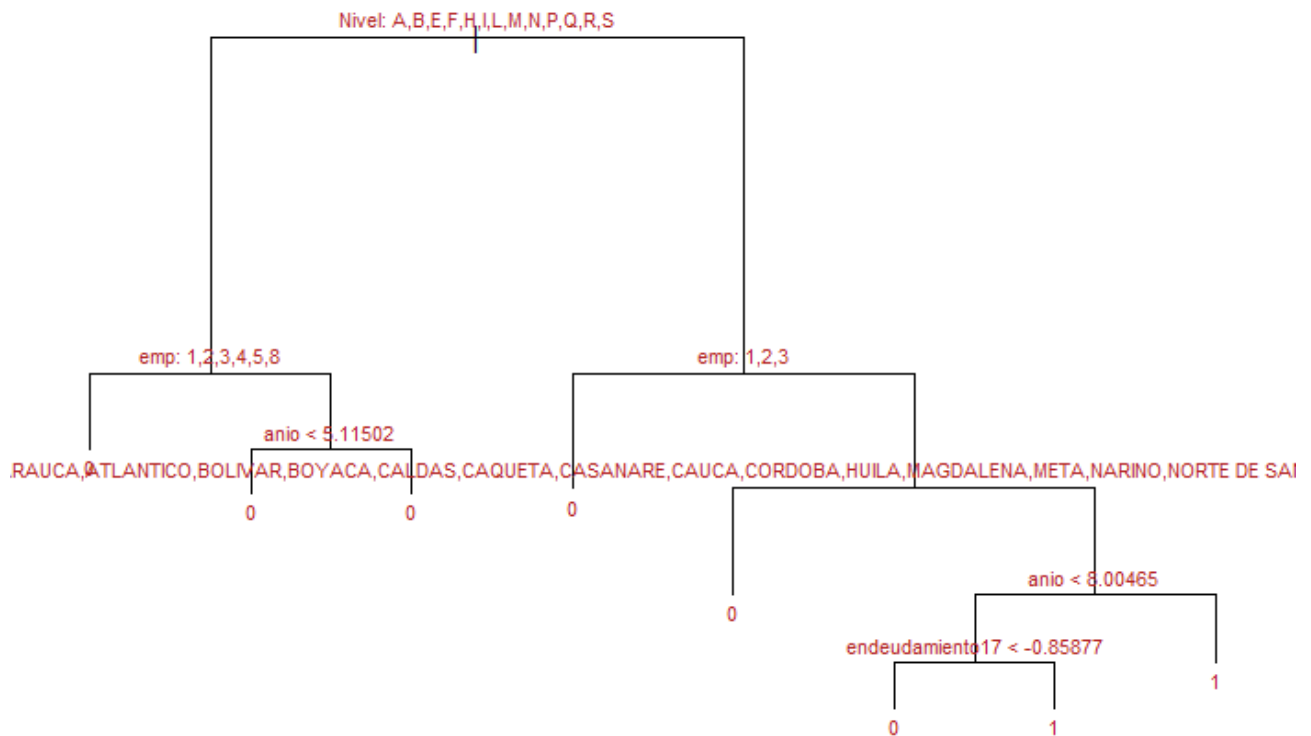


Figura 4.2: Árbol de selección generado a partir de la base nivelada

Al ver que esta base contiene buena información para generar modelos de clasificación, se utilizó para realizar un modelo de random forest para obtener un mejor modelo del que se puede lograr con un único árbol de decisión, para esto se seleccionaron los hiperparametros del número de árboles incluidos en el modelo (*num.trees*), profundidad máxima que pueden alcanzar los árboles (*max.depth*) y número de predictores considerados en cada división (*mtry*) a partir de una búsqueda por grilla, obteniendo una especificidad de 0.8917 logrando así un mejor score que el que se lograba alcanzar con un solo árbol.

Para tener una referencia de la capacidad del modelo anterior, Se realiza una predicción sobre toda la base, y así poder ver el comportamiento del modelo en situaciones reales de predicción, logrando con esta base conseguir un score superior al 81 % manteniendo su capacidad predictiva alta, teniendo estos valores como referencia, con el fin de lograr un modelo más ajustado se agregaron más variables, de una a la vez, para ver de que forma afectan al modelo. El primer ensayo se realizó con la variable cuantía de la base *coes*.

Cuantía: Un indicador del buen estado de una empresa es si esta mantiene relaciones comerciales con el estado, ya que esto le da una imagen de solides y seriedad que puede favorecerla

para llegar a ser una EAC. La primera variable de estudio dentro de esta base es la del valor total de las contrataciones con el estado, al realizar el cruce con la base solo 741 empresas tienen relaciones con el estado por lo que a las demás empresas se les asignó el valor de 0 para esta variable, para reducir la carga en el entrenamiento se discretizó la variable a partir de sus deciles. Para todos los ejercicios con las nuevas variables, se volvía a nivelar la base con el fin de que se tuvieran en cuenta los nuevos valores y también se volvía a realizar la búsqueda por grid para encontrar los mejores hiperparámetros antes del entrenamiento. En el caso de la cuantía se obtuvo como resultado una especificidad del 0.8729, lo cual si bien no es un mal valor, esta por debajo de los puntajes conseguidos anteriormente.

Contratos: Esta variable cuenta la cantidad de contratos que se tienen con el estado, se decidió analizarlas por separado ya que puede influir el hecho de que exista un solo contrato por una alta cuantía o varios contratos con cuantías pequeñas. La asignación de valores se realizó de la misma manera que en el caso anterior dejando como valor 0 para las empresas que no tienen relaciones con el estado y se discretizó creando categorías a partir de sus quintiles.

Incidentes judiciales: De la misma manera en que los contratos con el estado pueden generar una buena imagen empresarial, los incidentes judiciales generan una mala imagen y reflejan posibles malos manejos o problemas que a futuro podrían obstaculizar el ascenso de la empresa a ser una EAC. En la base se encontraron que 201 empresas registran algún tipo de incidencia, para ese caso dado que la cantidad de posibles incidentes judiciales es muy baja, se decidió solo transformar la variable a tipo factor (categórica) sin realizar ningún tipo de discretización adicional.

Modificación de estatutos: Esta variable puede reflejar las consecuencias, buenas o malas, causadas por los cambios en la estructura de la empresa, esta base solo cuantifica la cantidad de modificaciones realizadas por año, de las cuales posiblemente se encuentran alguna de las siguientes situaciones:

- Cambio de la razón social.
- Reforma de la vigencia.
- Reforma del objeto.
- Reforma del sistema de representación legal.
- Reforma a las facultades del representante legal.
- Reforma del quórum, forma y antelación de citación a las reuniones para los órganos colegiados (Junta de socios, asamblea de accionistas, miembros de juntas directivas, etc.).

- Transformación de sociedades.
- Conversión de empresas unipersonales.
- Fusión
- Escisión
- Otros

Así, a lo mas con esta base se captura que sucedió algún cambio en la empresa, pero no se puede especificar cual fue, al igual que en el caso anterior solo se transformó la variable en categoría. Para un ejercicio futuro seria bueno poder tener estos cambios desagregados, por ejemplo, tener el dato de fusiones puede ayudar mucho durante la preparación de la base para definir la supervivencia de una empresa.

Los resultados de cada modelo se pueden comparar en la siguiente tabla

Modelo	Especificidad
Árbol con datos sin balancear	0
Árbol con datos balanceados	0.8062
Random forest balanceado	0.8931
Predicción sobre toda la base	0.8147
Ensayo con la variable Cuantía	0.8674
Ensayo con la variable Contrato	0.8629
Ensayo con la variable Judiciales	0.8629
Ensayo con la variable Modificación de estatutos	0.862
Ensayo con todas las variables	0.862

Cuadro 4.1: *Resultados de los modelos con las nuevas variables.*

Random forest con las nuevas variables

Se realizo una prueba con todas las nuevas variables para identificar si actuando en conjunto se logra una mejor calidad en el modelo, como resultado se obtuvo un ajuste del 0.862 sin lograr un cambio significativo comparándolo con los modelos iniciales, mas aun al realizar una predicción sobre toda la base se obtuvo un score de 0.64, el mas bajo logrado durante todo el ejercicio.

Ensayo desagregando el CIIU

Se uso la variable de clasificación industrial a dos niveles, es decir la sección más el primer número correspondiente a la división, para ver si se logra una mejora el nivel de predicción, no se genera una desagregación más precisa ya que el sistema no soporta una alta cantidad de categorías en una misma variable y un alto nivel de especificidad podría hacer que se pierdan

relaciones generales. Con esta configuración Se obtuvo un valor de 0.8524, siendo inferior al conseguido con el CIIU a un solo nivel.

A parte del modelo anterior se realizo un árbol de clasificación y un random forest usando la base original (la que mejor resultado dio hasta el momento) mas la clasificación CIIU a dos niveles para ver si se lograban mejores resultados, pero de nuevo el resultado mostró que esta desagregación no ofrece ningún tipo de información que mejore el score obtenido.

4.0.1. Ensayos con otros modelos

Maquina de soporte vectorial

Para comprobar si el modelo de random forest es el mas adecuado tal como lo dice la literatura, se utilizo la base inicial, la que mejor score a generado para entrenar un modelo de support vector machine (svm). Para esto primero se realizó 10 cross-validation para identificar el valor óptimo de penalización.

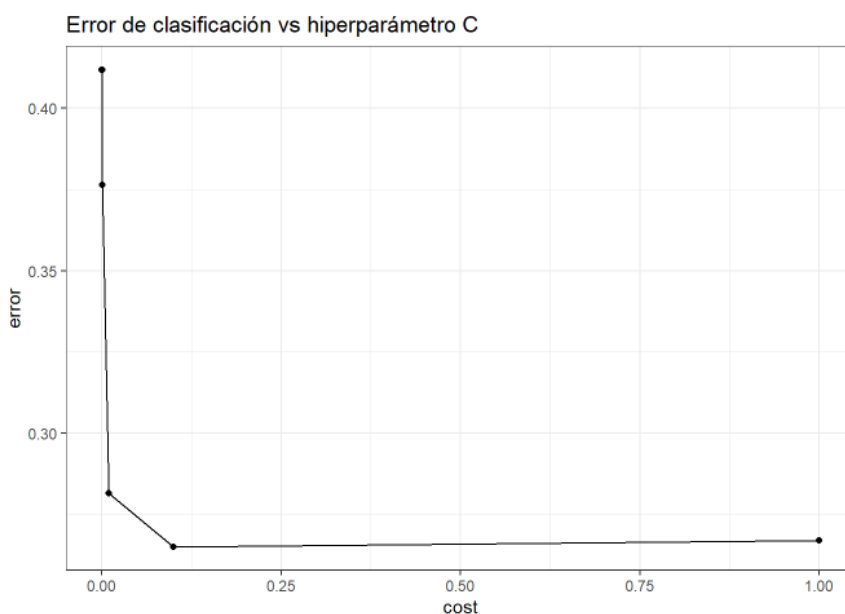


Figura 4.3: Resultado de la búsqueda del óptimo de penalización usando cross-validation

El proceso de cross-validation muestra que el valor de penalización con el que se consigue menor tasa error es de 0.10 o superior. Después de entrenar el modelo se consiguió un ajuste del 0.714 inferior al que se logro usando cualquier combinación de variables en el random forest.

Red neuronal

Tambien se entreno una red neuronal seleccionando sus hiperparámetros a partir de una búsqueda de hiperparametros por random grid search, el cual hace una búsqueda de combinaciones aleatorias en lugar de usar grid search cartesiano (todas las combinaciones posibles)

por ser poco práctico, además como algunas combinaciones aleatorias pueden ser muy poco favorables, se activó una parada temprana.

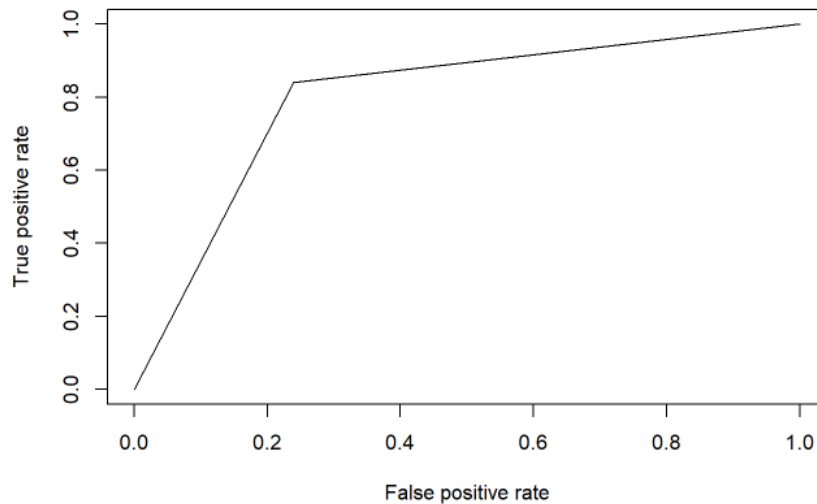


Figura 4.4: Curva ROC para el modelo de red neuronal

Con este modelo se obtuvo una exactitud del 0.7109, muy parecida a la del SVM he igualmente inferior a lo conseguido con los demás modelos.

4.0.2. Importancia de predictores

Al final de todos los ejercicios, el mejor modelo se consiguió con el random forest usando la base original. Para conocer mejor el peso de cada variable dentro de este modelo, se estudiaron los predictores que más influencia tienen dentro del modelo, para esta evaluación se usaran 2 métodos diferentes

Pureza de nodos

Este criterio se base en calcular el incremento total en la pureza de los nodos debido a las divisiones en las que participa el predictor que se esta evaluando, para esto en cada nodo se registra el descenso conseguido en la medida del Gini. Así para cada uno de los predictores se calcula el descenso medio conseguido en el conjunto de árboles, teniendo que cuanto mayor sea este valor, mayor la contribución de este predictor en el modelo. Por ejemplo, cuando se divide un nodo del árbol, el algoritmo busca un campo con la mejora más elevada del total de impureza, calculada como el total de impureza entre todos los nodos hijo restados del total de impureza del nodo padre.

Con este indicador se obtuvo que la cantidad de empleados, el nivel de endeudamiento y el la edad de la empresa son los valores que mas influyen en el entrenamiento del modelo.

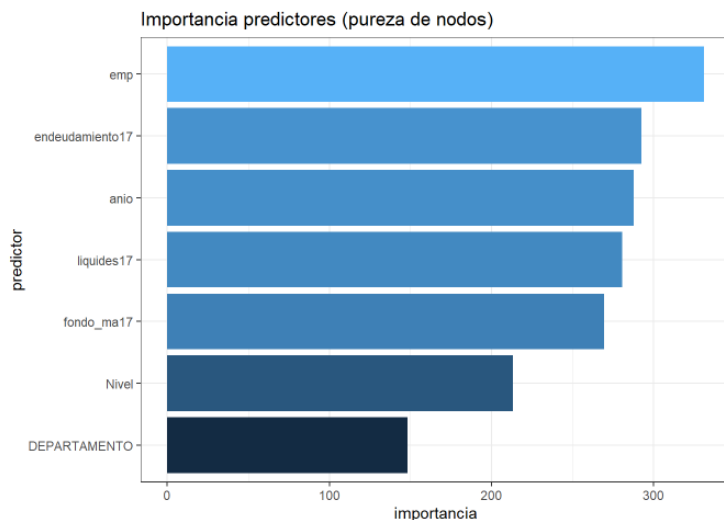


Figura 4.5: Importancia de los predictores usando el método de la pureza de nodos

Permutación

Este indicador consiste en crear un conjunto de árboles para luego permutar en todos ellos los valores del predictor que se está estudiando manteniendo los demás predictores constantes, luego de cada permutación se recalcula el incremento en el error debido a la permutación del predictor teniendo que si el predictor estaba contribuyendo al modelo, al perder la información que proporcionaba esa variable es de esperarse que el modelo aumente su error en un porcentaje que se puede interpretarse como la influencia que tiene el predictor sobre el modelo.

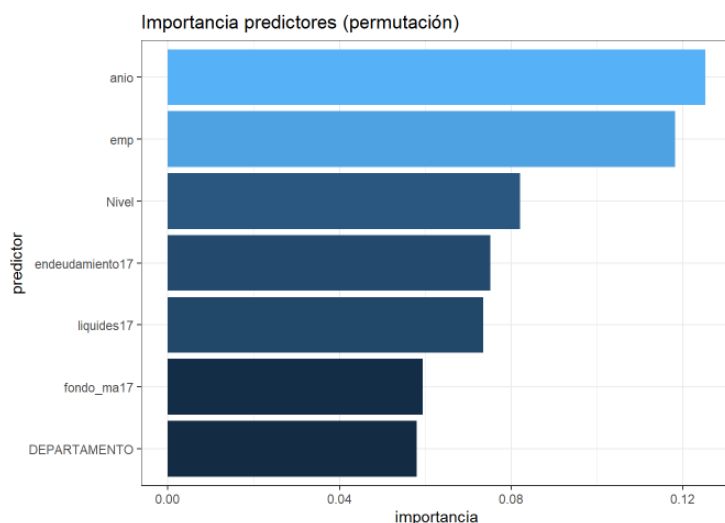


Figura 4.6: Importancia de los predictores usando el método de permutación

Aunque los niveles de influencia que muestran los diferentes métodos varia, se puede ver que ambos muestran a las mismas tres variables en las primeras posiciones, salvo el de permutación que agrega el predictor de *nivel* (Sector económico). Una de las cosas mas destacables de este ejercicio es ver como aparte de las variables que por estudios previos se esperaría que tuvieran una fuerte influencia en el modelo, se detecto que la variable de *endeudamiento* es muy relevante al clasificar a una empresa como EAC, esto posibilita que se realicen estudios posteriores sobre esta variable y su influencia en el ciclo de vida de una empresa.

Capítulo 5

Conclusiones

A partir de los resultados obtenidos en el desarrollo de este trabajo, se puede ver que en primer lugar si es posible entrenar modelos de aprendizaje automatizado con la suficiente exactitud para ser viables en la predicción de EAC, además se comprobó que, dentro de los diferentes modelos evaluados, el modelo de random forest ofrece mejor exactitud para este tipo de ejercicios. También se pudo corroborar que los predictores más importantes son la edad de la empresa favoreciendo a las empresas jóvenes, la cantidad de empleados, en donde se ve que las empresas de alto crecimiento están fuertemente ubicadas dentro de las pequeñas y medianas empresas. Por otro lado la variable endeudamiento no es una variable que se tenga en cuenta en estudios previos y que para este trabajo mostró ser relevante en la generación de los modelos lo cual abre nuevas posibilidades de enfoques en el estudio de EAC, también es curioso ver que, si bien la variable de ubicación por departamento influye en el entrenamiento del modelo, no lo hace con la fuerza esperada por lo que en ejercicios futuros sería interesante usar un nivel de desagregación mayor de la variable de ubicación

Para actividades futuras es posible repetir el ejercicio teniendo en cuenta las siguientes opciones:

- Ampliación de la cantidad de empresas usadas en el estudio, ya que las empresas vigiladas por la superintendencia de sociedades pueden presentar características particulares que no permitan una buena generalización del modelo para el análisis de cualquier empresa.
- Volver a realizar las pruebas usando las diferentes marcas que se generaron a partir de las diferentes metodologías de clasificación de EAC.
- Aumentar la especificidad del ejercicio al tratar de identificar empresas gacelas.
- Usar el modelo para generar predicciones sobre información proveniente de otro país para así evaluar el nivel de precisión del modelo con datos de otras economías.

- Aplicar otros modelos de aprendizaje automatizado, como por ejemplo el uso de deep learning para aprovechar su alta capacidad de clasificación a partir de la generación de modelos con diferentes topologías.
- Agregar información más variada, como puede ser geolocalización de las empresas a partir de sus coordenadas (obtenidos por registros administrativos, fuentes oficiales o web scraping de paginas con datos empresariales), o información de cambios estructurales de la empresa como fusiones o cambios de razón social.
- Realizar estudios locales a nivel de regiones o departamentos para ver si al analizar empresas presentes en un mismo entorno se logra una mejor predicción local.
- Uso de información de registros administrativos diferentes (por ejemplo la información proveniente de los pagos a seguridad social, que en el caso de Colombia son almacenados en la Planilla Integrada de Liquidación de Aportes - PILA) que permitan tener datos con un alto nivel de actualización y contenga información de relaciones como por ejemplo entre la empresa y sus empleados, o las relaciones entre empresas donde posiblemente existan factores que influyan en el desarrollo de una empresa para convertirse en EAC (es decir, ver como influye en el desarrollo de una empresa al tener relaciones con empresas de alto crecimiento y empresas comunes).

Bibliografía

- [1] *EUROSTAT - OECD manual on business demography statistics*. OECD, Paris [u.a.], 2007.
- [2] Guidelines on the use of statistical business registers for business demography and entrepreneurship statistics, 2018.
- [3] Falco J. Bargagli-Stoffi, Jan Niederreiter, and Massimo Riccaboni. Supervised learning for the prediction of firm dynamics. *In: Data Science for Economics and Finance (2021) Springer*, September 2020.
- [4] D.L. Birch, Massachusetts Institute Technology. Program on Neighborhood, and Regional Change. *The Job Generation Process*. MIT Program on Neighborhood and regional Change, 1984.
- [5] Alex Coad, Julian Frankish, Richard G. Roberts, and David J. Storey. Growth paths and survival chances: An application of gambler’s ruin theory. 28:615–632, 2013.
- [6] Arti Grover Goswami, Denis Medvedev, and Ellen Olafsen. High-growth firms: Facts, fiction, and policy options for emerging economies, 2019.
- [7] Mónica Ortiz Medina Julian Domínguez Rivera. *Manual de demografía empresarial adaptado para Colombia*. Confecamaras -Red de camaras de comercio, 2019.
- [8] Ciro Romero D Melgarejo Z and simon Elorz K. Empresas de alto crecimiento de colombia y españa. *Revista Dimensión Empresarial.*, 19(1), 2021.
- [9] The United Nations Committee of Experts on Business and Trade Statistics (UNCEBTS). *Methodological sheets for the Global core set of principal business and trade indicators*. United Nations.
- [10] Arjen van Witteloostuijn and Daan Kolkman. Is firm growth random? a machine learning perspective. 11:e00107, 2019.
- [11] Dongyuan Wu. Libreria rsbid, <https://github.com/dongyuanwu/RSBID>.

Capítulo 6

Anexos

6.1. Variables de los balances financieros

En el siguiente esquema se representa el nivel de dependencia entre las variables de los balances financieros, donde entre mas este a la derecha significa un nivel mayor de atomicidad junto con las formulas que permiten el calculo de variables mas generales.

1. R: Resultado del ejercicio
(RA-RIII)
 - RII: Impuesto de renta
(RAI-RAG)
 - RA: Resultados antes de impuestos
(RAI-RAG)
 - RAI: Total ingresos
(RAIE+RAIX)
 - RAIE Ingresos operacionales
(RAIE4120+RAIE4130)
 - ◇ RAIE4120: Ingresos de actividades ordinarias
 - ◇ RAIE4130: Otros ingresos operacionales
 - RAIX Ingresos no operacionales
 - RAG: TOTAL GASTOS
(RAGE+RAGX)
 - RAGE: Costos y Gastos operacionales
(RAGE51+RAGE52+RAGE55+RAGE60)
 - ◇ RAGE51: Gastos operacionales adm

- ◇ RAGE52: Gastos operacionales ventas
- ◇ RAGE55: Otros gastos operativos
- ◇ RAGE60: Costo de ventas
- RAGX: Gastos no operacionales

2. Tesorería:

(ACC-ACC114)/PSC

■ ACC: TOTAL ACTIVO CORRIENTE

(ACC211+ACC113+ACC114+ACC118+ACC11G+ACC11H)

- ACC211 Efectivo y equivalentes de efectivo
- ACC113 Cuentas por cobrar
- ACC114 Inventarios
- ACC118 Otros Activos
- ACC11G Activos biológicos
- ACC11H Activos por impuestos corrientes

■ ACC114: Inventarios

■ PSC : TOTAL PASIVO CORRIENTE

(PSC223+PSC224+PSC225+PSC126+PSC228+PSC12H)

- PSC223 Cuentas por Pagar
- PSC224 Pasivos por impuestos corrientes
- PSC225 Provisiones corrientes por beneficios a los empleados
- PSC126 Pasivos estimados y provisiones
- PSC228 Otros pasivos corrientes
- PSC12H Pasivos incluidos en grupos de activos para su disposición clasificados como mantenidos para la su disposición clasificados como mantenidos para la venta

3. Liquidez:

(ACL/ACC)

■ AC: ACTIVO

● ACL: TOTAL ACTIVO NO CORRIENTE

(ACL214+ACL115+ACL118+ACL11M+ACL11O+ACL11P+ACL11Q+ACL11R)

- ACL214 Inventarios no corrientes
- ACL115 Propiedad planta y equipo
- ACL118 Otros Activos

- ACL11M Propiedad de inversión
- ACL11O Activos biológicos no corrientes
- ACL11P Plusvalía
- ACL11Q Activos intangibles distintos de la plusvalía
- ACL11R Activos por impuestos diferidos
- ACC TOTAL ACTIVO CORRIENTE
(ACC211+ACC113+ACC114+ACC118+ACC11G+ACC11H)
 - ACC211 Efectivo y equivalentes de efectivo
 - ACC113 Cuentas por cobrar
 - ACC114 Inventarios
 - ACC118 Otros Activos
 - ACC11G Activos biológicos
 - ACC11H Activos por impuestos corrientes
- PSC: TOTAL PASIVO CORRIENTE

4. Solvencia:

(AC/PS)

- AC: ACTIVO
(ACL+ACC)
 - ACL TOTAL ACTIVO NO CORRIENTE
 - ACC TOTAL ACTIVO CORRIENTE
- PS: TOTAL PASIVO
(PSC+PSL)
 - PSC: TOTAL PASIVO CORRIENTE
 - PSL: TOTAL PASIVO NO CORRIENTE
(PSL224+PSL225+PSL228+PSL12J)
 - PSL224: Pasivos por impuestos corrientes
 - PSL225: Provisiones corrientes por beneficios a los empleados
 - PSL228: Otros pasivos corrientes
 - PSL12J: Pasivo por impuestos diferidos

5. Endeudamiento:

(PS/PT)

- PS: TOTAL PASIVO
(PSC+PSL)

- PT: PATRIMONIO

(PTT131+PTT133+PTT136+PTT237)

- PSL224: Pasivos por impuestos corrientes
- PSL225: Provisiones corrientes por beneficios a los empleados
- PSL228: Otros pasivos corrientes
- PSL12J: Pasivo por impuestos diferidos

6. Rentabilidad:

(R/RAIE)

- R: Resultado del ejercicio
- RAIE: Ingresos operacionales

7. Fondo de maniobra:

(AC-PC)

- AC: Activo
- PSC: TOTAL PASIVO CORRIENTE

8. Ratio de Cobertura de Intereses:

(RA/RAGXFI)

- RA: Resultados antes de impuestos
- RAGXFI: Gastos financieros