Universidad Nacional Mayor de San Marcos
Universidad del Perú. Decana de América

Dirección General de Estudios de Posgrado
Facultad de Ingeniería de Sistemas e Informática
Unidad de Posgrado

Modelo predictivo del éxito de las Startups de tecnología de la información usando Machine Learning

TESIS

Para optar el Grado Académico de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software

AUTOR

Edilberto VÁSQUEZ CIEZA

ASESOR

Dr. David Santos MAURICIO SÁNCHEZ

Lima, Perú

2023

Referencia bibliográfica

Vásquez, E. (2023). *Modelo predictivo del éxito de las Startups de tecnología de la información usando Machine Learning*. [Tesis de maestría, Universidad Nacional Mayor de San Marcos, Facultad de Ingeniería de Sistemas e Informática/Unidad de Posgrado]. Repositorio institucional Cybertesis UNMSM.

Metadatos complementarios

Datos de autor	
Nombres y apellidos	Edilberto Vásquez Cieza
Tipo de documento de identidad	DNI
Número de documento de identidad	43545322
URL de ORCID	https://orcid.org/0000-0002-7111-000X
Datos de asesor	
Nombres y apellidos	David Santos Mauricio Sánchez
Tipo de documento de identidad	DNI
Número de documento de identidad	06445495
URL de ORCID	https://orcid.org/0000-0001-9262-626X
Datos del jurado	
Presidente del jurado	
Nombres y apellidos	Cayo Víctor León Fernández
Tipo de documento	DNI
Número de documento de identidad	07001405
Miembro del jurado 1	
Nombres y apellidos	Fany Yexenia Sobero Rodríguez
Tipo de documento	DNI
Número de documento de identidad	20120467
Miembro del jurado 2	
Nombres y apellidos	Félix Armando Fermín Pérez
Tipo de documento	DNI
Número de documento de identidad	08736347
Datos de investigación	
Línea de investigación	C.0.3.22. Ingeniería de software

Grupo de investigación	Inteligencia Artificial - INTGARTI
Agencia de financiamiento	Perú. Universidad Nacional Mayor de San Marcos. Vicerrectorado de Investigación y Posgrado. Proyecto de investigación para grupos. C19200841
Ubicación geográfica de la investigación	País: Perú Departamento: Lima Provincia: Lima Distrito: Lima Latitud: -12.053268 Longitud: -77.085577
Año o rango de años en que se realizó la investigación	2018 - 2022
URL de disciplinas OCDE	Ingeniería de sistemas y comunicaciones https://purl.org/pe-repo/ocde/ford#2.02.04 Informática y Ciencias de la Información https://purl.org/pe-repo/ocde/ford#1.02.00

UNIVERSIDAD NACIONAL MAYOR DE SAN MARCOS
Universidad del Perú. Decana de América
Facultad de Ingeniería de Sistemas e Informática
Vicedecanato de Investigación y Posgrado
Unidad de Posgrado

ACTA DE SUSTENTACIÓN DE TESIS PARA OPTAR EL GRADO ACADÉMICO DE MAGÍSTER EN INGENIERÍA DE SISTEMAS E INFORMÁTICA CON MENCIÓN EN INGENIERÍA DE SOFTWARE

A los cuatro (04) días del mes de diciembre de 2023, siendo las 11:00 am., se reunieron en el Auditorio, Profesor: Alfredo Celso Alva Bravo, el Jurado de Tesis conformado por los siguientes docentes:

Dr. Cayo Víctor León Fernández (Presidente)
Mg. Fany Yexenia Sobero Rodriguez (Miembro)
Mg. Félix Armando Fermín Pérez (Miembro)
Dr. David Santos Mauricio Sánchez (Miembro Asesor)

Se inició la Sustentación invitando al candidato a Magíster ***EDILBERTO VÁSQUEZ CIEZA,*** *para que realice la exposición oral de la tesis para optar el Grado Académico de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software, siendo la Tesis intitulada:*

"MODELO PREDICTIVO DEL ÉXITO DE LAS STARTUPS DE TECNOLOGÍA DE LA INFORMACIÓN USANDO MACHINE LEARNING"

Concluida la exposición, los miembros del Jurado de Tesis procedieron a formular sus preguntas que fueron absueltas por el graduando; acto seguido se procedió a la evaluación correspondiente, habiendo obtenido la siguiente calificación:

DIECISIETE (17) MUY BUENO

Por tanto, el presidente del Jurado, de acuerdo con el Reglamento General de Estudios de Posgrado, otorga al Bachiller ***EDILBERTO VÁSQUEZ CIEZA*** *el Grado Académico de Magíster en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software.*

Siendo las. ..... horas, el presidente del Jurado de Tesis, da por concluido el acto académico de Sustentación de Tesis.

Dr. Cayo Víctor León Fernández
(Presidente)

Mg. Fany Yexenia Sobero Rodríguez
(Miembro)

Mg. Félix Armando Fermín Pérez
(Miembro)

Dr. David Santos Mauricio Sánchez
(Miembro Asesor)

CERTIFICADO DE SIMILITUD

Yo David Santos Mauricio Sánchez en mi condición de asesor acreditado con Dictamen N° 000316-2023-UPG-VDIP-FISI/UNMSM de la tesis de investigación, cuyo título es "Modelo predictivo del éxito de las Startups de tecnología de la información usando Machine Learning", presentado por el egresado Edilberto Vásquez Cieza para optar el grado de Magister en Ingeniería de Sistemas e Informática con mención en Ingeniería de Software, CERTIFICO que se ha cumplido con lo establecido en la Directiva de Originalidad y de Similitud de Trabajos Académicos, de Investigación y Producción Intelectual. Según la revisión, análisis y evaluación mediante el software de similitud textual, el documento evaluado cuenta con el porcentaje de 19% de similitud, nivel **PERMITIDO** para continuar con los trámites correspondientes y para su **publicación en el** repositorio institucional.

Se emite el presente certificado en cumplimiento de lo establecido en las normas vigentes, como uno de los requisitos para la obtención del grado/ título/ especialidad correspondiente.

Firma del Asesor _____

DNI: 06445495

Nombres y apellidos del asesor: David Santos Mauricio Sánchez

DEDICATORIA

A mis padres, por los valores inculcados y apoyo brindado.

A mis hermanos, por su constante soporte para mi formación profesional.

AGRADECIMIENTOS

A mi asesor Dr. David Santos Mauricio Sánchez, por su constante apoyo, orientación y revisiones para la elaboración del presente trabajo de investigación.

A mis padres, por su apoyo constante en todo el proceso de mi educación y por ser el soporte en mi vida.

A mis hermanos, por su permanente apoyo para lograr mis objetivos personales y profesionales.

A mis profesores de maestría de la Universidad Nacional Mayor de San Marcos, por compartir sus conocimientos durante los dos años de estudio.

Al jurado, por su revisión, aportes y comentarios sobre el trabajo realizado como mejora continua.

A la Universidad Nacional Mayor de San Marcos, por el financiamiento parcial de la investigación.

TABLA DE CONTENIDOS

LISTA DE TABLAS viii

RESUMEN xv

Palabras clave: Startups, factores críticos de éxito, pronóstico, machine learning xv

abstract xvi

1.1. Antecedentes del problema 17
1.2. Problema 18
1.3. Motivación 19
1.4. Objetivos 20
1.4.1. Objetivo general 20
1.4.2. Objetivos específicos 20
1.5. Propuesta 20
1.6. Organización de la tesis 21

CAPÍTULO 2: MARCO TEÓRICO 23

2.1. Las Startups 23
2.1.1. Concepto de Startup 23
2.1.2. Ciclo de vida de las Startups 23
2.1.3. Concepto de éxito 24
2.1.4. Factores críticos de éxito 25
2.2. Modelos predictivos basados en Machine Learning 25

CAPÍTULO 3: ESTADO DEL ARTE 31

3.1. Metodología 31
3.2. Planificación 32
3.2.1. Preguntas de investigación 32
3.2.2. Bancos de datos 33
3.2.3. Palabras claves 33
3.2.4. Criterios de inclusión y exclusión 34
3.3. Desarrollo 34
3.3.1. Proceso de búsqueda 34
3.3.2. Selección de estudios 35
3.4. Resultados 36
3.4.1. Análisis 37
3.4.2. Factores Críticos de Éxito de las Startups (Q1) 38
3.4.3. Modelos de predicción del éxito de las Startups (Q2) 41
3.4.4. Herramientas para predecir el éxito de las Startups (Q3) 42
3.4.5. Discusión 43

CAPÍTULO 3: MÉTODO Y MODELO PREDICTIVO DEL ÉXITO DE LAS STARTUPS DE TI 45

3.1. Método para generar modelos predictivos del éxito de Startups de TI 45
3.1.1. Selección de factores 46
3.1.2. Extracción de datos 46
3.1.3. Preprocesamiento 47
3.1.4. Aprendizaje 47
3.2. Modelo de predicción del éxito de Startups de TI 48
3.3. Implementación del modelo 51

3.3.1. Elicitación de requisitos 51
3.3.2. Casos de uso 54
3.3.3. Diseño 55
3.3.4. Arquitectura 66
3.3.5. Funcionamiento 69

CAPÍTULO 4: VALIDACIÓN 72
4.1. El dataset 72
4.2. Métricas 72
4.3. Preprocesamiento 73
4.4. Selección de factores 75
4.4.1. Análisis de componentes principales 76
4.4.2. Greedy Step Wise 78
4.5. Resultados 80
4.6. Análisis y discusión 87

CAPÍTULO 5: CONCLUSIONES, LIMITACIONES Y TRABAJOS FUTUROS 89
5.1. Conclusiones 89
5.2. Limitaciones 90
5.3. Trabajos futuros 91

REFERENCIAS 92

ANEXOS 101

A. Autovectores generados en la selección de factores con PCA 101
B. Matriz de correlación generada en la selección de factores con PCA 102
C. Factores que influyen en el éxito de una Startup de TI 103
D. Producción científica de la investigación 108

LISTA DE TABLAS

Tabla 1. *Cadenas de búsqueda aplicados a los campos de los bancos de datos* 33

Tabla 2. *Criterios de inclusión y exclusión de artículos* 34

Tabla 3. *Artículos potenciales y seleccionados por banco de datos* 36

Tabla 4. *Factores críticos de éxito de las Startups identificados y usados en la predicción* 38

Tabla 5. *Modelos y técnicas usados para predecir el éxito de las Startups* 42

Tabla 6. *Herramienta de software basado en técnicas de Machine Learning para predecir el éxito de las Startups* 43

Tabla 7. *Requisitos funcionales para el desarrollo del software* 51

Tabla 8. *Requisitos no funcionales para el desarrollo del software* 53

Tabla 9. *Atributos de la Entidad Usuarios* 56

Tabla 10. *Atributos de la Entidad Rol* 57

Tabla 11. *Atributos de la Entidad Startup* 57

Tabla 12. *Atributos de la Entidad Predicciones* 58

Tabla 13. *Atributos de la Entidad Factores* 58

Tabla 14. *Atributos de la Entidad Modelo* 59

Tabla 15. *Métricas usadas en el proceso de predicción de las Startups de TI* 72

Tabla 16. *Factores y clases de atributos seleccionados* 74

Tabla 17. *Características del dataset original y preprocesado* 75

Tabla 18. *Ranking de atributos generados a través de PCA* 76

Tabla 19. *Factores seleccionados por el método Greedy Step Wise* 79

Tabla 20. *Resultados de la validación cruzada para el escenario de 20 factores con 10-fold* 80

Tabla 21. *Resultados de la validación cruzada para el escenario de 17 factores con 10-fold* 81

Tabla 22. *Resultados de la validación cruzada para el escenario de 5 factores con 10-fold* 81

Tabla 23. *Matriz de confusión de los resultados de testing para los tres escenarios* 82

Tabla 24. *Resultados de pronóstico del éxito de un STI para 7 modelos de ML y los 3 híbridos en los escenarios de 23 factores y 20 factores.* 84

Tabla 25. *Resultados de pronóstico del éxito de un STI para 7 modelos de ML y los 3 híbridos en escenarios de 17 factores y 5 factores* 85

LISTA DE FIGURAS

Figura 1. Método para generar modelo de ML de pronóstico del éxito de una Startup de TI 21
Figura 2. Ciclo de vida de las Startups 24
Figura 3. Representación del clasificador Naive Bayes 25
Figura 4. Representación de un árbol de decisión 26
Figura 5. Representación del algoritmo Random Forest 27
Figura 6. Representación del clasificador Gradient Boosting como suma de *n* árboles de decisión 28
Figura 7. Representación del algoritmo SVM 28
Figura 8. Representación de una Red Neuronal Artificial 29
Figura 9. Representación del clasificador K-NN 30
Figura 10. Modelo de revisión de literatura de Kitchenham & Charters (2007) 32
Figura 11. Flujo de proceso de revisión de literatura 35
Figura 12. Artículos por banco de datos 36
Figura 13. Artículos por año de publicación 37
Figura 14. Método para generar modelo de Machine Learning de predicción del éxito de una Startup 46
Figura 15. Diagrama del proceso de aprendizaje del modelo de Machine Learning 48
Figura 16. Modelo para la predicción del éxito de una Startup de TI 49
Figura 17. Modelo híbrido de machine learning de predicción del éxito de una Startup 50
Figura 18. Diagrama de casos de uso del sistema 54
Figura 19. Diagrama de clases del sistema 55
Figura 20. Diagrama de Entidad-Relación del sistema 60
Figura 21. Wireframe de interfaz de usuario de Login 61
Figura 22. Wireframe de interfaz de usuario de dashboard 62
Figura 23. Wireframe de interfaz de usuario Registro de Startup 63
Figura 24. Wireframe de interfaz de usuario de Predicción (selección de Startups y modelo) 64
Figura 25. Wireframe de interfaz de usuario de Predicción (Registro de factores) 65
Figura 26. Wireframe de interfaz de usuario de Predicción (Procesamiento) 65
Figura 27. Wireframe de interfaz de usuario de Resultados de Predicción 66

Figura 28. Arquitectura de referencia propuesta del sistema 67
Figura 29. Diagrama de despliegue del sistema 68
Figura 30. Diagrama de despliegue del sistema 69
Figura 31. Diagrama de despliegue del sistema 70
Figura 32. Diagrama de despliegue del sistema 71
Figura 33. Variación de la pérdida por época del modelo Perceptron Multicapa para el pronóstico del éxito de un STI: a) primer escenario; b) segundo escenario; c) tercer escenario; d) cuarto escenario. 84

RESUMEN

El pronóstico del éxito de una Startup de Tecnología de la Información (STI) es un problema muy complejo debido a diversos factores e incertidumbre que la afecta. El enfoque del aprendizaje automático (ML) es prometedor porque presenta buenos resultados para problemas de pronóstico, sin embargo, presenta diversidad de parámetros, procesos, factores y datos que requieren ser considerados para mejorar los resultados del pronóstico. En este trabajo, se propone un método sistemático para construir un modelo predictivo de éxito de una Startup de TI con alta precisión, basado en los factores que influyen en el éxito y algoritmos de ML. El método consta de 4 procesos, un modelo híbrido y un inventario de 79 factores críticos de éxito, además, es aplicable a cualquier ciudad o región. El método fue aplicado a una base de datos de 265 Startups de TI de Australia con 7 algoritmos de ML (SVM, Perceptron Muli-layer, Decision Tree, Naive Bayes, KNN, Radom Forest y Gradient Boosting) y 3 modelos híbridos basados en la estrategia de Votación, asimismo, el algoritmo GreedyStepWise para reducir los factores. Los resultados muestran que el método permite obtener modelos de pronóstico con mejores resultados para los algoritmos, en promedio, incrementa la precisión en 11.69 %, la especificidad en 3.25 % y la exactitud en 21.75 %, pero, en general, el modelo híbrido proporciona mejores resultados alcanzando una exactitud ideal del 100 %. GreedyStepWise permite identificar 5 factores (size startup, company revenue, r&d, financial capital and global economic enviroment) con los que se obtiene, a través de SVM y los modelos híbridos, pronóstico con precisión del 82 % y exactitud del 88 %.

Palabras clave: Startups, factores críticos de éxito, pronóstico, machine learning

ABSTRACT

Predicting the success of a Startup in Information Technology (SIT) is a very complex problem due to the diverse factors and uncertainty that affect it. The focus of machine learning (ML) is promising because it presents good results for prediction issues; however, it presents a diversity of parameters, factors, and data that require consideration to improve prediction results. In this study, a systematic method is proposed to build a predictive model for SIT success, based on factors. The method consists of four processes, a hybrid model, and an inventory of 79 success factors. The method was applied to a database of 265 SITs from Australia with seven ML algorithms and three hybrid models based on the Voting strategy and the GreedyStepwise algorithm to reduce the factors. On average, precision increments in 11.69%, specificity in 3.25%, and accuracy in 21.75%; the prediction has precision of 82% and accuracy of 88%.

Keywords: Startups, critical success factors, forecast, machine learning

CAPÍTULO 1: INTRODUCCIÓN

1.1. Antecedentes del problema

La innovación se define como la introducción de un nuevo o significativamente mejorado producto (bien o servicio), de un proceso, de un nuevo método de comercialización o de una forma de organización, en las prácticas internas de la empresa, la organización de lugar de trabajo o las relaciones exteriores (Manual de Oslo OCDE, 2006).

En cuanto al ecosistema de innovación, este está compuesto principalmente por dos agentes. En primer lugar, se tiene a los proveedores de recursos que pueden ser entidades de Gobierno o a través de inversores privados como capitales de riesgo (Venture Capital), inversionistas ángeles y otros, dispuestos a invertir en ideas innovadoras, compartiendo el riesgo en la fase inicial, y que, de acuerdo con el éxito que tenga, pueden recuperar rápidamente su inversión. En segundo lugar, se puede mencionar a los innovadores, que pueden ser empresas, asociaciones o personas agrupadas alrededor de una idea innovadora de negocio, estos últimos conocidos como Startup.

El análisis del éxito de una Startup se hace a través de la identificación de los Factores Críticos de Éxito que pueden ser tanto internos como externos a la organización, considerándose, por ejemplo, (i) Atributos de la Startup, (ii) Atributos de los miembros, (iii) Atributos del entorno, entre otros (Kakadiya, 2015).

Con el advenimiento de nuevas tecnologías y el uso de modelos, se vienen planteando herramientas basadas en técnicas de inteligencia artificial para predecir el éxito de las Startups, tomando como insumos a los Factores Críticos de Éxito (CSFs). Entre estas técnicas, se tiene, por ejemplo, el uso de Data Mining y Machine Learning.

1.2. Problema

El éxito puede definirse de diversas formas, dependiendo del sector donde se esté tratando, así, por ejemplo, en relación con las Startups, se puede definir desde el punto de vista del emprendedor, del inversionista o del usuario. En particular, para el emprendedor, esto puede significar el incremento de sus ingresos, realización personal, de igual modo, para los inversionistas, puede significar el incremento de sus utilidades. Mas aun teniendo en cuenta que las Startups – emprendimientos con base tecnológica– tienen fases definidas durante su ciclo de vida, también se puede definir el éxito por cada etapa.

Tomando en consideración las características antes descritas, el problema de la presente tesis es la baja tasa de precisión en la predicción del éxito de las Startups en todas sus fases de desarrollo.

Importancia del problema

En la industria de la Startup de TI, las altas tasas de natalidad van de la mano con un gran riesgo de fracaso, puesto que solo una de cada tres sobrevive los primeros tres años, por lo que, desafortunadamente, la tasa de fracaso de este tipo de empresas es alto en todo el mundo (Pugliese et al., 2016). De acuerdo con Ejermo & Xiao (2014), entre los años 1990 y 2000, solo el 21 % de las Startups de TI en Suecia sobrevivieron después de 5 años. Por otro lado, Hyder & Lussier (2016) afirma que más del 80 por ciento de las Startups fracasan en su primer año de existencia.

El interés por restringir el foco en la Startup de TI recae en el hecho de que este tipo de empresas representa una parte importante en el crecimiento de las economías de los países en vías de desarrollo (Banda y Lussier, 2015). De acuerdo con Thanh (2015), la ratio de fracaso de la Startup llega a ser del 75 %.

En el Perú, según el Ministerio de la Producción, en el año 2015, el número de empresas que sobrevive después de 5 años de vida está por debajo del 50 % y, en particular, la ratio de fracaso de la Startup es del 70 %, demostrando que cada vez es más difícil lanzar y mantener una iniciativa tecnológica a lo largo del tiempo. Desde el 2013, el Ministerio de la Producción, a través de su programa StartUp Perú, ha beneficiado a 309 Startups que provienen de 18 regiones, de las cuales el 92 % son de TI, invirtiendo más de 20 millones de soles (PRODUCE, 2017).

1.3. Motivación

A pesar de la importancia que tiene la innovación tecnológica en la competitividad de las empresas y la dinamización de los mercados, así como de los pocos estudios teóricos existentes en relación con la predicción del éxito de las Startups de TI a nivel mundial, se tiene lo siguiente:

- Los modelos predictivos actuales no contemplan la predicción en las etapas durante el ciclo de vida de una Startup.
- Los pocos modelos predictivos que existen en la literatura no involucran más de 14 factores, habiendo más factores relevantes que influyen en el éxito de las Startups de TI.
- La tasa de precisión aún sigue siendo baja (<88 %)

En este escenario, se hace necesario el desarrollo de un modelo de predicción del éxito de las Startups basándose en técnicas y herramientas de Machine Learning, que permita a los innovadores identificar, evaluar y predecir el éxito de sus emprendimientos y a los inversionistas (Programas de financiamiento público o Capitales de riesgo privados) a invertir en proyectos exitosos, lo que redundará en un menor riesgo para sus inversiones.

1.4. Objetivos

1.4.1. Objetivo general

Diseñar un método y un modelo predictivo del éxito de las Startups de TI, basado en técnicas de Machine Learning que permita predecir el éxito de las Startups de TI con tasas de precisión superior a los modelos existentes.

1.4.2. Objetivos específicos

i. Identificar los Factores Críticos de Éxito que influyen en cada etapa del proceso de innovación.
ii. Determinar los requisitos de las Startups de TI por fases de su ciclo de vida.
iii. Diseñar un método sistemático para generar modelos predictivos del éxito de las Startups de TI basado en técnicas de Machine Learning que contemple los Factores y requisitos durante las fases de su clico de vida.
iv. Construir una herramienta de predicción del éxito de las Startups de TI.

1.5. Propuesta

Un método y un modelo predictivo del éxito de Startups basado en técnicas de Machine Learning, que permita predecir el éxito de las Startups de TI teniendo como datos de entrada a los factores críticos de éxito que influyen en cada etapa del su ciclo de vida. En la Figura 1 se muestra el esquema del método propuesto.

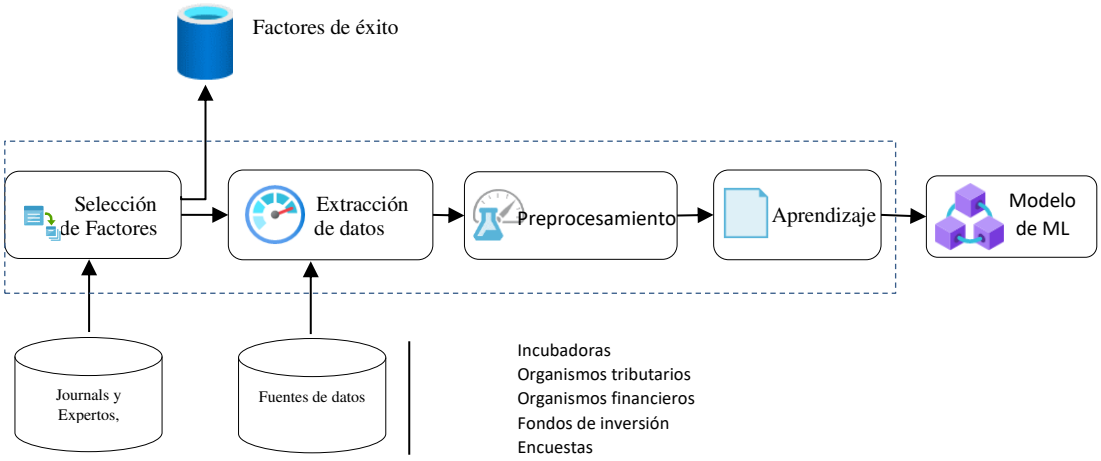

Figura 1. Método para generar modelo de ML de pronóstico del éxito de una Startup de TI

1.6. Organización de la tesis

El presente trabajo de investigación está organizado en cinco capítulos.

En el capítulo 1, se hace una introducción del trabajo de investigación, donde se indica los antecedentes del problema, su definición e importancia, motivación, objetivos y resumen de la propuesta planteada.

En el capítulo 2, se desarrolla el estado del arte con relación a factores críticos de éxito, técnicas de predicción, herramientas de predicción del éxito de las Startups, para ello se sigue un esquema de Metodología, Planificación, Desarrollo y Análisis de la información.

En el capítulo 3, se plantea un método y un modelo de predicción del éxito de las Startups de TI como aporte, de este modo, se detalla sus componentes, técnicas y herramientas utilizadas.

En el capítulo 4, se describen las actividades y tareas de validación, así como los resultados obtenidos.

En el capítulo 5, se presentan las conclusiones del trabajo de investigación y los trabajos a futuro.

CAPÍTULO 2: MARCO TEÓRICO

En este capítulo, se definen algunos conceptos principales sobre el tema de investigación. Inicia con una sección relacionada a las Startups, en el cual se detalla su concepto, ciclo de vida y el éxito. Luego se describe las técnicas de Machine Learning existentes para hacer análisis predictivos, de este modo, se mostrará las 7 técnicas que se usarán en el presente trabajo.

2.1. Las Startups

2.1.1. Concepto de Startup

Las Startups son organizaciones compuestas por innovadores cuya misión es desarrollar un producto o servicios exitosos y que tienen como fuente primaria una idea innovadora. Así mismo, Nadežda P., et al., (2019) define a la Startup como una agrupación de personas alrededor de una idea innovadora con base tecnológica y con un modelo de negocio replicable y escalable, sus características se mencionan a continuación:

- Organizaciones en fases iniciales de constitución
- Asociados a la innovación o tecnología
- Proyección de crecimiento

2.1.2. Ciclo de vida de las Startups

Las Startups, durante su ciclo de vida, pasan por etapas o fases. Según (OCDE & Eurostat, 2006), se define 4 etapas: la etapa seed, donde no existe un plan de negocios definido al 100 %, y el equipo de trabajo es pequeño y deben de encargarse de darle forma a dicho plan. Además, se suele recurrir al capital semilla o, lo que es lo mismo, a las aportaciones de los fundadores, sus familiares o algún pequeño inversor;

la etapa early, donde el producto ya está en el mercado y cada día hay más clientes dispuestos a comprarlo, por eso es necesario seguir innovándolo; la etapa growth, momento en que el modelo de negocio planteado en la fase inicial se ha perfeccionado, lo que provoca la aparición de fondos de inversión especializados en la financiación de la Startup; por último, la etapa de expansión, fase en la que se establecen alianzas con otras empresas para facilitar el asentamiento en otros mercados. En la Figura 2, se muestra las fases de ciclo de vida de una Startup.

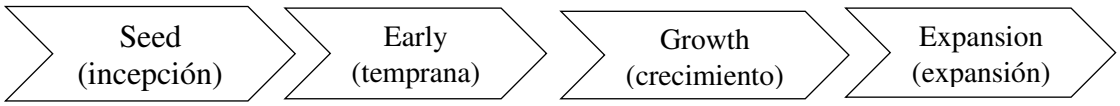

Figura 2. Ciclo de vida de las Startups

Fuente: Elaboración propia

2.1.3. Concepto de éxito

De los estudios seleccionados, hay algunos que intentan definir el éxito de una Startup. Por ejemplo, de acuerdo con Martens D. et al., (2011), el éxito de esta se debe al crecimiento en las ventas y obtener una buena rentabilidad. Además, en el estudio de Elhedhli S. et al., (2014), se considera al éxito como el buen desempeño financiero de la empresa. Por otro lado, Santisteban & Mauricio (2017) definen a una Startup exitosa como aquella nueva empresa que ofrece productos y/o servicios innovadores y que genera nuevos puestos de trabajos especializados.

2.1.4. Factores críticos de éxito

Los factores críticos de éxito son variables que condicionan el éxito o fracaso de las Startups y su importancia radica en que pueden influir de forma positiva o negativa en las diferentes etapas del ciclo de su ciclo de vida.

2.2. Modelos predictivos basados en Machine Learning

2.3.1. Concepto

Los modelos predictivos son representaciones conceptuales de un conjunto de factores relacionados con el éxito y técnicas computacionales que permiten estimar anticipadamente el éxito (pronóstico) de los emprendimientos (personas naturales o empresas en edad temprana) con base tecnológica en TI.

2.3.2. Modelos predictivos

2.3.2.1. Naive Bayes

Es el modelo predictivo basado en redes bayesianas, el mismo que se basa en probabilidades, donde se asume que la presencia (o ausencia) de una determinada característica de una clase no está relacionada con la presencia (o ausencia) de cualquier otra característica.

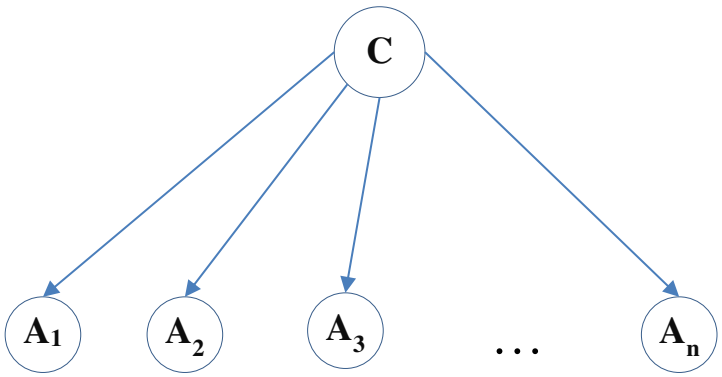

Figura 3. Representación del clasificador Naive Bayes

Fuente: Elaboración propia

2.3.2.2. Árbol de decisión

Es definido como el algoritmo de predicción, clasificación y regresión de tipo supervisado, el cual utiliza atributos discretos y continuos. Se compone de un nodo de decisión (raíz), nodos de probabilidades (ramas) y nodos de salida (hojas). En la Figura 4, se muestra los elementos de un árbol de decisión.

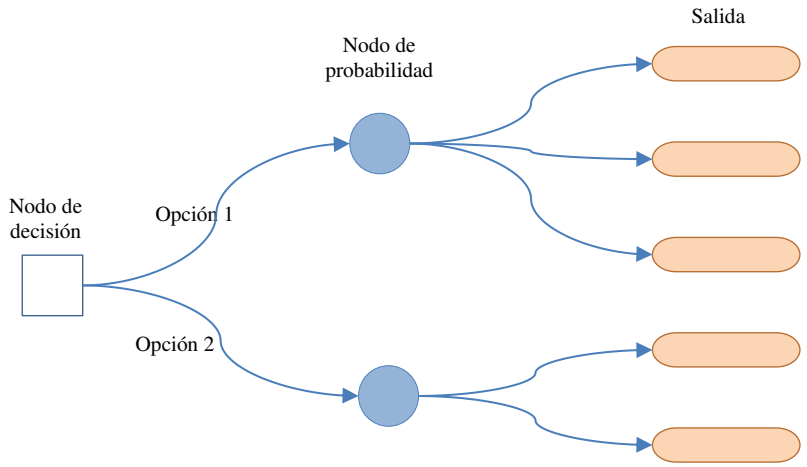

Figura 4. Representación de un árbol de decisión

Fuente: Elaboración propia

2.3.2.3. Random Forest

El clasificador Random Forest es un modelo de aprendizaje de tipo supervisado desarrollado por Breiman, L. (2001) el cual consta de múltiples árboles de decisión generados a partir de un conjunto de datos de entrenamiento. La última instancia de una clase en un modelo Random Forest se asigna mediante la salida de la clase que es el nodo de las salidas. En la Figura 5, se muestra la representación del clasificador Random Forest.

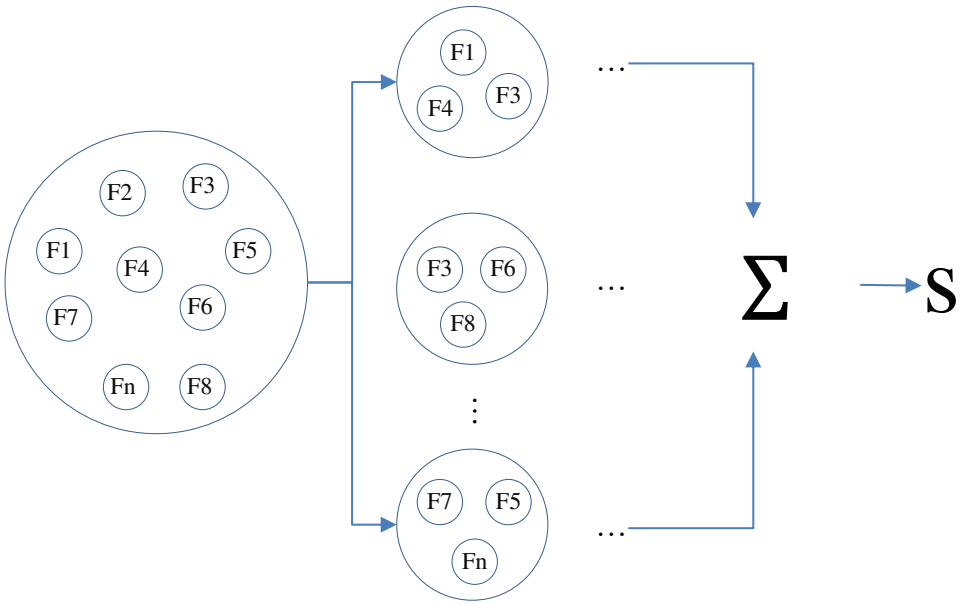

Figura 5. Representación del algoritmo Random Forest

Fuente: Elaboración propia

2.3.2.6. Gradient Boosting

Es un modelo desarrollado por Friedman, J. H. (2001) basado en el trabajo de Freund,Y. & Schapire, R. E. (1997) quienes propusieron un modelo denominado AdaBooster basado en árboles de decisión. Gradient Boosting es una mejora de AdaBooster agregando una función denominada Gradiente, que permite minimizar el error a medida que se agregan más árboles al modelo.

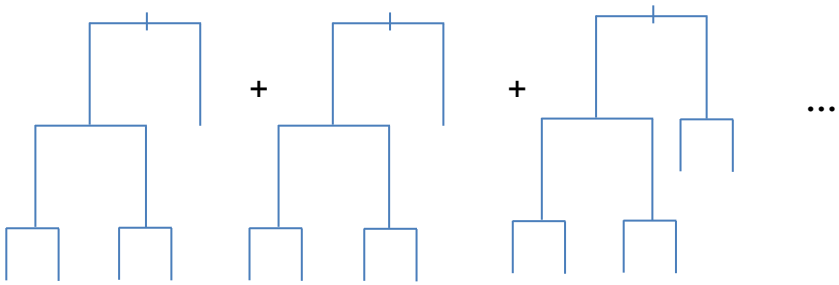

Figura 6. Representación del clasificador Gradient Boosting como suma de *n* árboles de decisión

Fuente: Elaboración propia

2.3.2.4. Support Vector Machine

Se trata de un algoritmo de clasificación supervisado que permite optimizar una función de transformación para distinguir entre clases de dos instancias diferentes maximizando el margen alrededor de un hiperplano que separa las dos clases. La función de transformación es denominada función Kernel y es la encargada de incrementar la dimensionalidad a fin de separar las clases de un conjunto de datos separable.

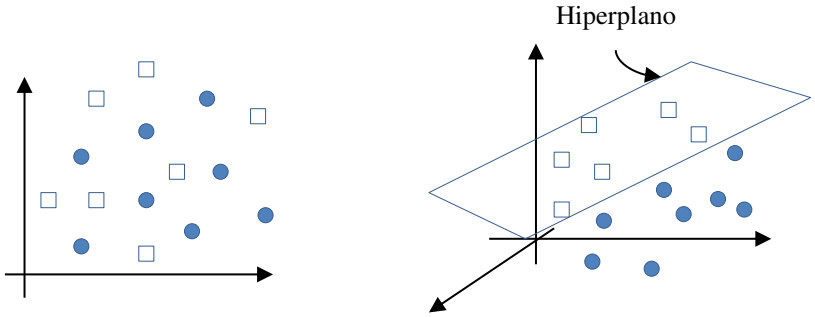

Figura 7. Representación del algoritmo SVM

Fuente: Elaboración propia

2.3.2.4. Redes Neuronales

Consisten en nodos funcionales que están organizados en capas las cuales realizan una suma ponderada y la transformación de sus entradas.

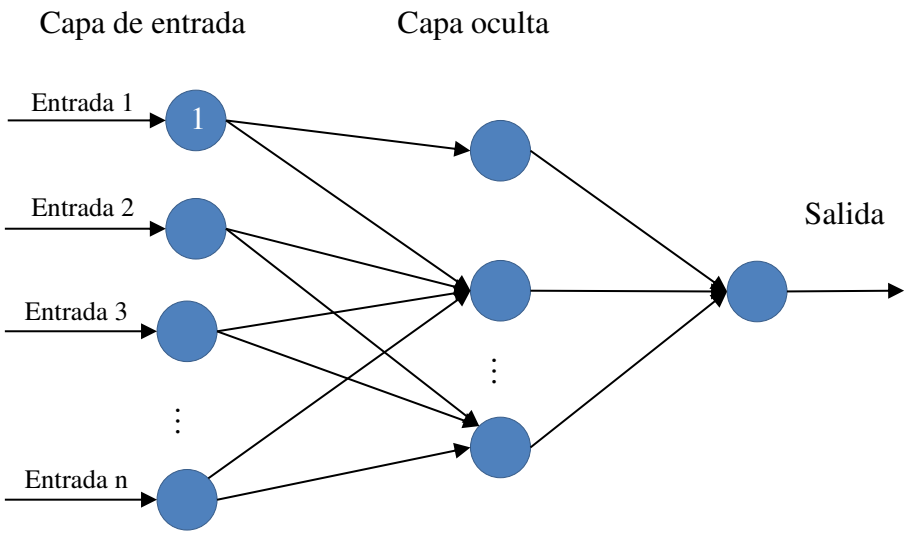

Figura 8. Representación de una Red Neuronal Artificial

Fuente: Elaboración propia

2.3.2.7. Karest-Nearest Neighbor

Es un algoritmo de clasificación de tipo supervisado que consiste en asignar un elemento dado a una clase o conjunto de entrenamiento en función a la distancia de dicho elemento con los del conjunto. El número de vecinos más cercanos al elemento a predecir es denotado por "*k*". En la Figura 9, se muestra la representación del algoritmo K-NN.

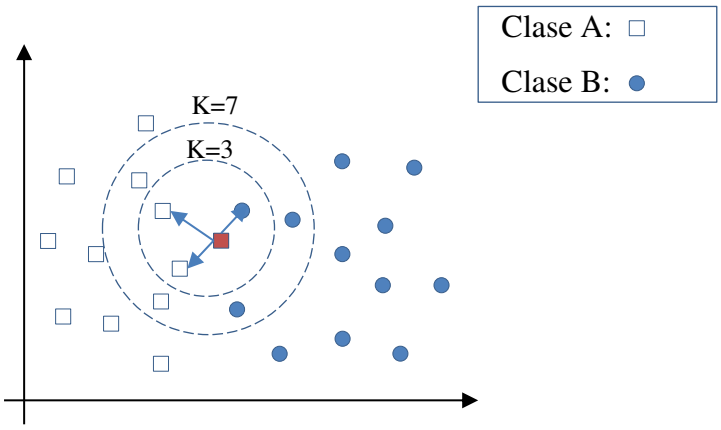

Figura 9. Representación del clasificador K-NN

Fuente: Elaboración propia

CAPÍTULO 3: ESTADO DEL ARTE

En este capítulo se muestra el estado del arte de las investigaciones relacionadas con los Factores Críticos de Éxito de las Startups, los Modelos Predictivos del éxito, y las herramientas que se utilizan para predecir el éxito de las Startups, donde se toma en cuenta el ámbito de alcance a nivel mundial. Este capítulo está organizado en cuatro secciones: (1) Metodología de la revisión de literatura, (2) Planificación de la revisión de literatura, (3) Desarrollo de la revisión de literatura y (4) Resultados y análisis de la revisión de literatura.

3.1. Metodología

Se realizó una revisión de la literatura considerando una metodología de tres fases usadas para este tipo de revisión propuesto por Kitchenham & Charters. (2007) y referenciadas en trabajos de similares características por Santisteban & Mauricio (2017), así mismo, por Cabrera & Mauricio (2017), adaptada para la presente investigación y organizada en tres fases que se mencionan a continuación:

- ✓ *Planificación:* En esta fase, se elaboran las preguntas de investigación y se define el protocolo de revisión.
- ✓ *Desarrollo:* En esta etapa, se seleccionan los estudios primarios de acuerdo con los criterios de selección y exclusión.
- ✓ *Resultados:* En este paso, se presentan las estadísticas y el análisis realizado a los estudios previamente seleccionados.

En cada una de las fases, se desarrollan actividades orientadas a lograr una revisión sistemática de literatura que puede ser adaptada para cada dominio.

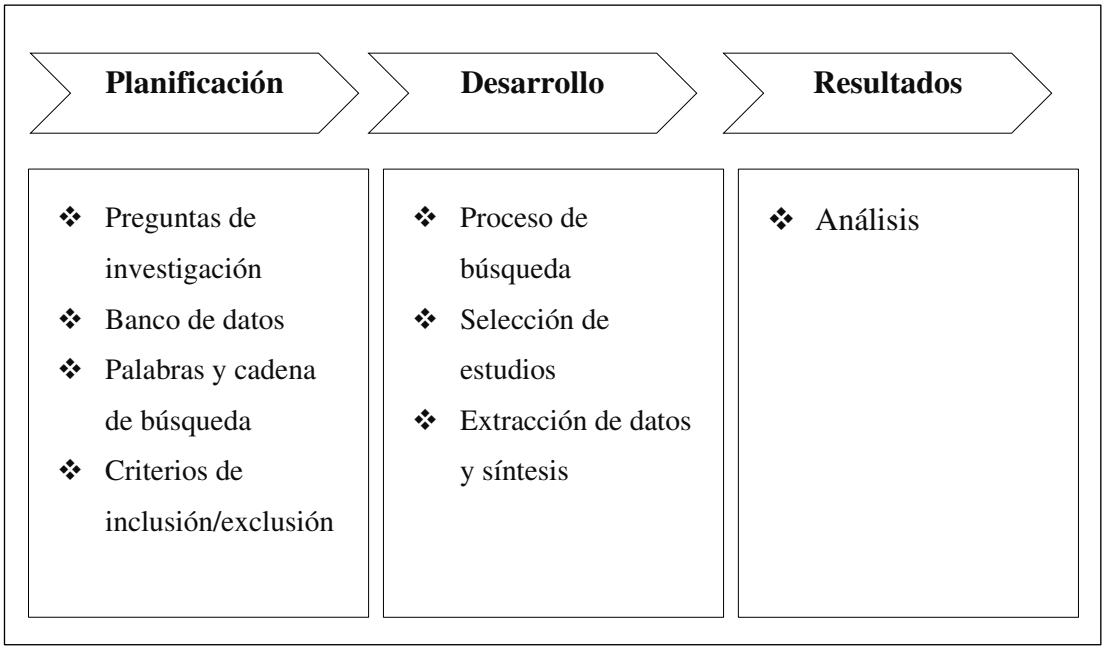

Figura 10. Modelo de revisión de literatura de Kitchenham & Charters (2007)

Fuente: Elaboración propia

En la Figura 10, se muestra el modelo de revisión de literatura planteado por Kitchnham el cual está compuesto por tres fases y sus actividades.

3.2. Planificación

En esta fase, se ha definido los elementos claves como son las preguntas relacionadas a la investigación, los bancos de datos, las palabras clave (ver Tabla 1) y los criterios de inclusión y/o exclusión (ver Tabla 2) a fin de obtener artículos relevantes para el trabajo.

3.2.1. Preguntas de investigación

Para lograr el objetivo del capítulo, se han planteado tres preguntas de investigación que están directamente relacionadas con el tema de la tesis, las cuales se mencionan a continuación:

Q1. ¿Cuáles son los factores críticos de éxito de las Startups de TI?

Q2. ¿Qué métodos, modelos y técnicas de predicción del éxito de las Startups de TI existen?

Q3. ¿Qué herramientas de software de predicción del éxito de las Startups existen?

3.2.2. Bancos de datos

- SCIENCE DIRECT
- SPRINGER
- ACM
- IEEE EXPLORER
- EMERALD INSIGHT
- OTROS

3.2.3. Palabras claves

Las cadenas construidas para la búsqueda son las siguientes: (*Critical Success Factors*) OR (*Startup*), (*Prediction model*) OR (*Startup*); (*Prediction model*) OR (*Innovation Projects*) aplicados a los campos de Título, Resumen y Palabras clave.

Tabla 1. *Cadenas de búsqueda aplicados a los campos de los bancos de datos*

Cadena de búsqueda	Campo de búsqueda
(Critical Success Factors) OR (*Startup*)	Title, Abstract and Keywords
(Prediction Model) OR (*Startup*)	Title, Abstract and Keywords
(Prediction Model) OR (*Innovation project)*	Title, Abstract and Keywords

Fuente: Elaboración propia

3.2.4. Criterios de inclusión y exclusión

Los criterios de inclusión y exclusión para el proceso de revisión de literatura para el presente trabajo de investigación de muestran en la Tabla 2.

Tabla 2. *Criterios de inclusión y exclusión de artículos*

Criterios de inclusión	Criterios de exclusión
Periodo: 2014-2022	Artículos en idioma distinto al inglés
Artículos de Journal con factor de impacto SJR	
Artículos relacionados con las preguntas de investigación	
Artículos con aporte y validación	

Fuente: Elaboración propia

3.3. Desarrollo

Esta fase consiste en la ejecución de la metodología de búsqueda de información y contempla la definición del proceso de búsqueda de información, la selección de estudios y la revisión y síntesis de los trabajos seleccionados.

3.3.1. Proceso de búsqueda

Para la fase de ejecución o desarrollo de la metodología de búsqueda de información relevante para la investigación, se ha tomado como base los ítems definidos en la fase de planificación considerando los criterios de inclusión y exclusión establecidos en la Tabla 2. El flujo del proceso de búsqueda de artículos se muestra en la Figura 11.

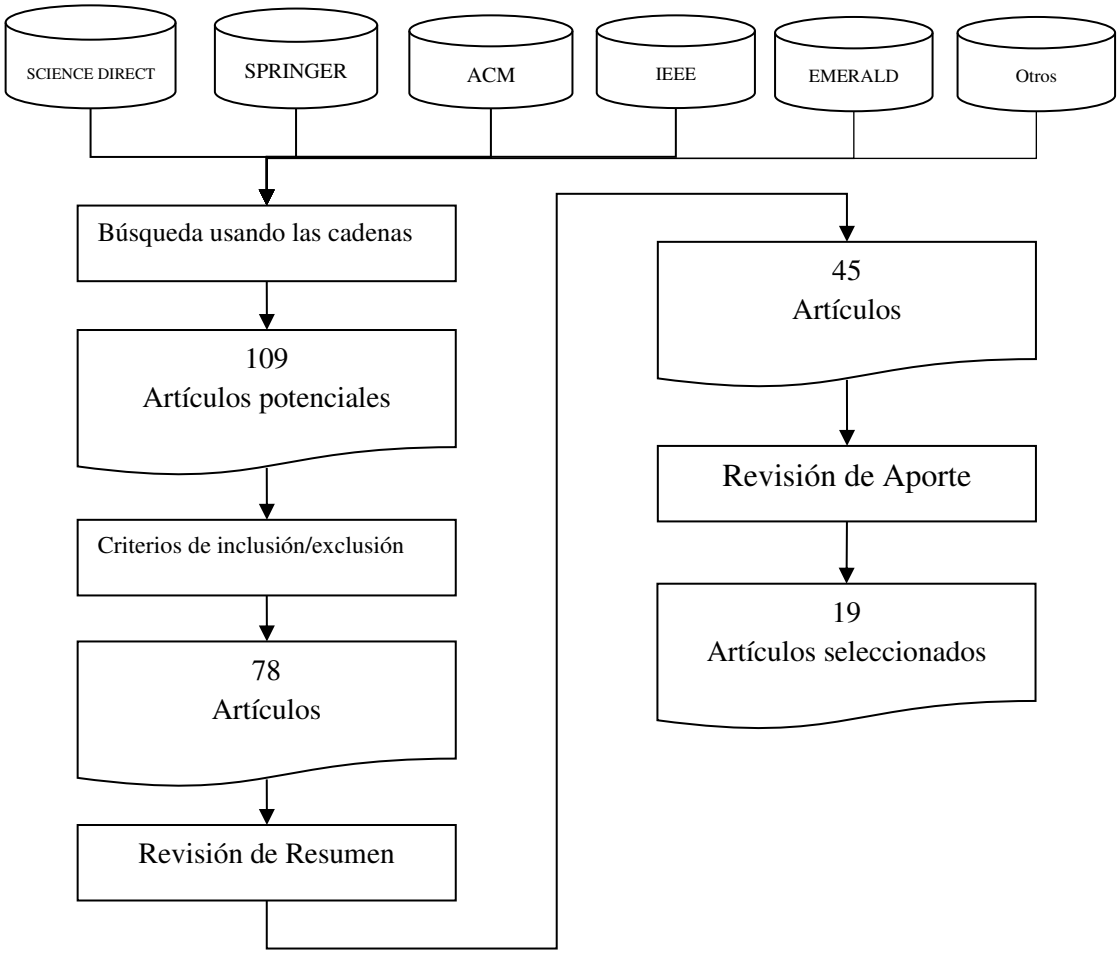

Figura 11. Flujo de proceso de revisión de literatura

Fuente: Elaboración propia

3.3.2. Selección de estudios

Después del desarrollo de las actividades de búsqueda y revisión de artículos en los bancos de datos Science Direct, Springer, ACM y IEEE Explore, se obtuvo 109 artículos candidatos. Luego del refinamiento de la búsqueda basados en los criterios de exclusión, se obtuvo 78 artículos, para finalmente hacer una revisión del resumen, aportes y validación, obteniendo 19 artículos, resultados que se muestran en la Tabla 3.

3.4. Resultados

Tabla 3. *Artículos potenciales y seleccionados por banco de datos*

Banco de datos	Artículos potenciales	Artículos seleccionados
Science Direct	24	10
Springer	10	2
ACM	6	2
IEEE Explore	3	1
Emerald	11	3
Otros	5	1
Total	109	19

Fuente: Elaboración propia

En el proceso de búsqueda de artículos en los diferentes bancos de datos, se obtuvo una mayor proporción en Science Direct, seguido por Emerald, Springer y ACM IEEE Explore y otros, tal como se muestra en la Figura 12.

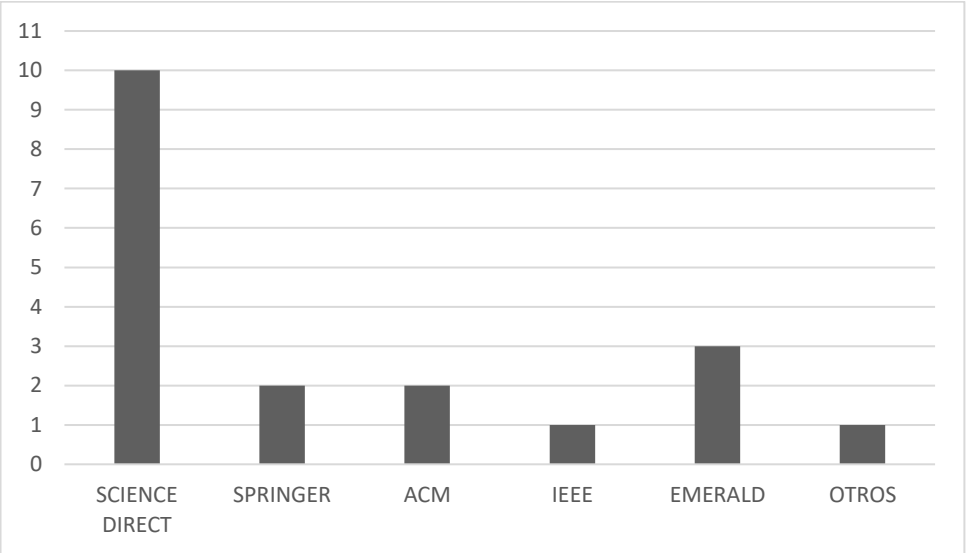

Figura 12. Artículos por banco de datos

Fuente: Elaboración propia

En cuanto a la tendencia de los artículos por año de publicación, se tiene que la mayor proporción de ellos se encuentran en el año 2015, seguido por el año 2016, 2014, 2017 y, finalmente, el año 2018, tal como se muestra en la Figura 13.

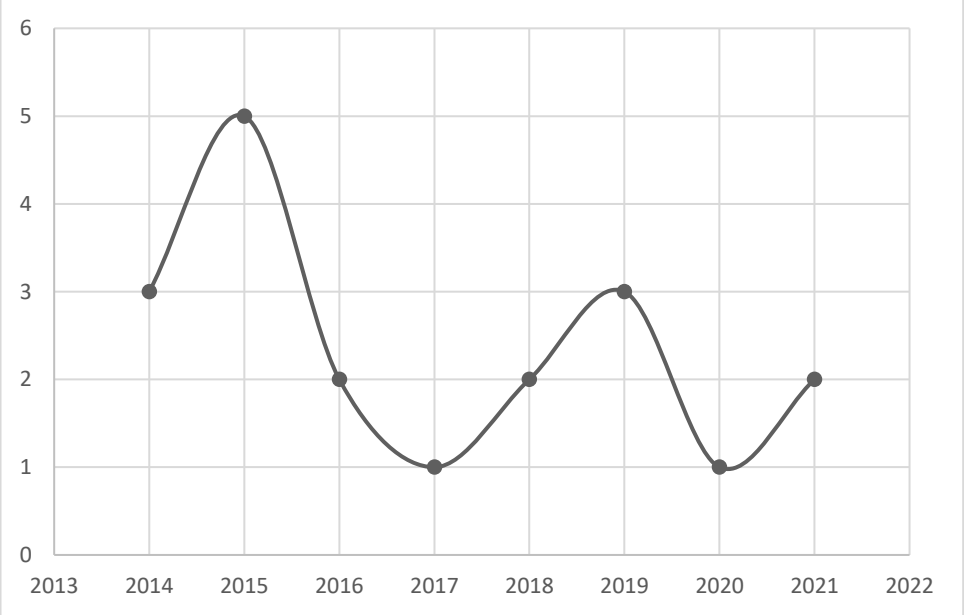

Figura 13. Artículos por año de publicación

Fuente: Elaboración propia

3.4.1. Análisis

Luego de la revisión de los artículos seleccionados aplicando la metodología de la sección anterior, se detalla los resultados obtenidos a través de una lectura exhaustiva, lo que permite responder a las preguntas de investigación planteadas.

3.4.2. Factores Críticos de Éxito de las Startups (Q1)

Con la finalidad de conocer los factores críticos de éxito, analizar su incidencia, así como sus relaciones, es que desde el año 1984 se vienen haciendo estudios. Por ejemplo, en el trabajo de Van de Ven et al. (1984), se menciona la importancia de la motivación de los emprendedores como un factor clave. Por su parte, Álvarez (2001) analiza a la tecnología como un factor importante en la dinámica de las Startups; Hormiga, et al. (2010) agrega a la ubicación como factor, entendiendo que, mientras más cerca de un polo de desarrollo está, se tiene mayores probabilidades de éxito.

En (Santisteban & Mauricio, 2017), se hace una revisión de literatura y se identifican 21 factores, adicionalmente, 10 factores se proponen en (Santisteban et al., 2021a), y en (Santisteban et al., 2021b) se estudia la influencia de 27 factores sobre el ciclo de vida de una Startup de TI.

Por otro lado, los Factores Críticos de Éxito también son usados en la predicción del éxito de las Startups, en este sentido algunos autores han desarrollado modelos que hacen uso de los mismos, así, por ejemplo, Helabí & Lussier (2014) desarrollan un modelo de predicción para el caso de Chile, donde usan 8 factores; también Kakadiya et al. (2015), en su modelo predictivo del éxito de startups, usa 10 factores categorizados como Factores asociados a la compañía, Factores asociados al equipo y Factores del entorno; por otro lado, Krishna et al. (2016), en su modelo desarrollado, hace uso de 8 factores para predecir el éxito; finalmente, Tomy & Pardede (2018) utiliza 23 factores críticos de éxito en su modelo de predicción.

Tabla 4. *Factores críticos de éxito de las Startups identificados y usados en la predicción*

Id	Factor	Referencia de identificación	Referencia de uso en predicción

F1	Capital de trabajo	(Bocken, 2015; Grilli & Murtinu, 2014; Almakenzi et al., 2015; Bertoni et al., 2011; Colombo et al., 2010; Kim & Heshmati, 2010; Strehle et al., 2010; Yoon-Jun, 2010)	Helabi & Luissier (2014); Krishna et al. (2016); Tomy & Pardede (2018)
F2	Tecnología	(Alvarez & Barney, 2001)	Helabi & Luissier (2014); Tomy & Pardede (2018)
F3	Experiencia en gestión	(Groenewegen & De Langen, 2012; Van Gelderen et al., 2005; Anh et al., 2012; Arruda et al., 2013; Baptista et al., 2007; Bou-Wen et al., 2006; Cannone & Ughetto, 2014; Hyder & Lussier, 2016; Strehle et al., 2010; Thiranagama & Edirisinghe, 2015; Yoo et al., 2012; Fini et al., 2009)	Helabi & Luissier (2014); Kakadiya et al. (2015
F4	Planeamiento		Helabi & Luissier (2014
F5	Educación	(Van Gelderen et al., 2005; Baptista et al., 2007; Bou-Wen et al., 2006; Colombo et al., 2004; Dautzenberg & Reger, 2010; Davis & Zweig, 2005; Gartner & Liao, 2012; Hyder & Lussier, 2016; Pugliese et al., 2016; Rojas & Huergo, 2016; Thiranagama & Edirisinghe, 2015)	Helabi & Luissier (2014); Kakadiya et al. (2015)
F6	Partners	(Sefiani & Bown, 2013)	Helabi & Luissier (2014); Tomy & Pardede (2018)
F7	Marketing	(Dimov, Sheperd, & Sutcliffe, 2007)	Helabi & Luissier (2014); Tomy & Pardede (2018)
F8	Competencia	(Song, Podoynitsyna, Van der Bij, & Halman, 2008)	Kakadiya et al. (2015); Tomy & Pardede (2018)
F9	Años en el mercado	(Haltiwanger et al., 2012)	Kakadiya et al. (2015); Krishna et al. (2016)
F10	Ubicación	(Hormiga, Batista-Canino, & Sánchez-Medina, 2010)	Kakadiya et al. (2015)
F11	Innovación en producto	(Almus & Nerlinher, 1999)	Tomy & Pardede (2018)

F12	Clustering	(Maine et al., 2010; Yoon-Jun, 2010; Mueller et al., 2012)	Tomy & Pardede (2018)
F13	Política de ciencia y tecnología	Scarborough & Zimmerer, 2003)	Tomy & Pardede (2018)
F14	Dinamismo del entorno	Timmons & spinelli, 2004)	Tomy & Pardede (2018)
F15	Tamaño organizacional	(Song et al., 2008; Ganotakis, 2012; Baptista et al., 2007; Bou-Wen et al., 2006; Colombo et al., 2004; Dautzenberg & Reger, 2010; Gartner & Liao, 2012; Rojas & Huergo, 2016; Thiranagama & Edirisinghe, 2015; Gottschalk & Niefert, 2013; Joshi & Satyanarayana, 2014; Cannone & Ughetto, 2014; Strehle et al., 2010)	Tomy & Pardede (2018)
F16	Apoyo de gobierno	Lasch et al., 2007; Chorev & Anderson, 2006; Anh et al., 2012; Arruda et al., 2013; Davis & Zweig, 2005; Pugliese et al., 2016)	Tomy & Pardede (2018)
F17	Motivación inicial del emprendedor	(Greve & Salaff, 2003; Reynolds & Miller, 1992)	-
F18	Edad del emprendedor	(Oakey, 2003)	-
F19	Genero del emprendedor	(Becchetti & Trovato, 2002)	-
F20	Liderazgo del emprendedor	(Schneider et al., 2007; Wei-Wen, 2009)	-
F21	Experiencia en I & D del equipo fundador	Baum & Silverman, 2004)	-
F22	Capacidades tecnológicas / empresariales del equipo fundador	Garcia-Muiña & Navas-López, 2007; Groenewegen & De Langen, 2012; Yoon-Jun, 2010; Li et al., 2010)	-
F23	Experiencia previa puesta en marcha del	(Van Gelderen et al., 2005; Song et al., 2008; Baptista et al., 2007; Bou-Wen et al., 2006; Colombo et al., 2004; Dautzenberg & Reger, 2010; Davis & Zweig, 2005; Friar & Meyer,	-

	equipo fundador	2003; Gartner & Liao, 2012; Kim & Heshmati, 2010; Pugliese et al., 2016; Mueller et al., 2012; Bocken, 2015)	
F24	Experiencia en la industria del equipo fundador	(Spyros & Nickolaos, 2012; Preisendorfer et al., 2012; Anh et al., 2012; Baptista et al., 2007; Bou-Wen et al., 2006; Colombo et al., 2004; Dautzenberg & Reger, 2010; Friar & Meyer, 2003; Gartner & Liao, 2012; Hyder & Lussier, 2016; O'Regan & Sims, 2008; Pugliese et al., 2016; Rojas & Huergo, 2016; Thiranagama & Edirisinghe, 2015; Wei-Wen, 2009; Yoo et al., 2012)	-

Fuente: Elaboración propia

En la literatura, diversos autores definen factores de forma similar, por lo que pueden ser agrupados y una lista completa de 79 factores de muestra, tal como se observa en el Anexo C.

3.4.3. Modelos de predicción del éxito de las Startups (Q2)

La predicción del éxito de las Startups se hace a través de modelos, los mismos que usan técnicas de Machine Learning. Así, por ejemplo, Helabí & Lussier (2014) proponen un modelo probabilístico para determinar el éxito o fracaso de pequeñas empresas; Kakadiya et al. (2015) propone un modelo para el caso de India, donde utiliza dos técnicas que son árboles de decisión y el algoritmo basado en reglas; Krishna et al. (2016), en su intención de analizar el comportamiento de las Startups de la India, propone otro modelo de predicción del éxito basado en factores y 6 técnicas avanzadas de Machine que son analizados de forma individual; Tomy & Pardede (2018) analizan el caso de Australia y proponen un modelo de predicción considerando 12 factores y tres técnicas que son SVN, Naive bayes y K-NN. En la tabla 5, se muestran algunas técnicas.

Tabla 5. *Modelos y técnicas usados para predecir el éxito de las Startups*

Técnica	Referencia
Árboles de decisiones	Kakadiya et al. (2015); Krishna et al. (2016); Cerpa et al. (2016)
Algoritmos basados en reglas	Kakadiya et al. (2015)
Naive Bayes	Krishna et al (2016); Tomy & Pardede (2018); Cerpa et al. (2016)
Regresión Logística	Krishna et al. (2016); Cerpa et al. (2016)
Lazylb1	Krishna et al. (2016), Martens et al. (2011)
Random Forest	Krishna et al. (2016); Cerpa et al. (2016
Redes neuronales	Cerpa et al. (2016)
Redes Bayesianas	Krishna et al. (2016), Akhavan et al. (2021)
Support Vector Machine (SVM)	Tomy & Pardede (2018); Cerpa et al. (2016)
K-NN	(Tomy & Pardede, 2018); Cerpa et al. (2016)
Probit	(Helabí & Lussier, 2014)
XGBost	Ross et al (2021)
Text Datamining	Antretter et al. (2018)

Fuente: Elaboración propia

3.4.4. Herramientas para predecir el éxito de las Startups (Q3)

Krishna, A., et al., (2015) y Kakadiya (2015) implementaron sus modelos de predicción del éxito de las Startups a través de un Framework de Minería de Datos y Machine Learning llamado “WEKA”, que fue desarrollado por la Universidad Waikato de Nueva Zelanda. Además, otros trabajos como Yankov, B. & Vintanov, N., (2016), en su trabajo sobre herramientas para la predicción del éxito de las Startups en

Bulgaria, identifican dos herramientas desarrolladas para este propósito que son “Odds of Success Calculator” y “Blueprint start-up success calculator”.

Tabla 6. *Herramienta de software basado en técnicas de Machine Learning para predecir el éxito de las Startups*

Herramientas	Descripción	Referencia
WEKA	Plataforma de software para análisis en técnicas de Machine Learning y Minería de Datos escrito en Java y desarrollado por la Universidad de Waikato de Nueva Zelanda	Kakadiya, S., Krishna, A.
Odds of Success Calculator	Aplicación web desarrollada para calcular las probabilidades de éxitos de una compañía o Startup	Yankov, B.
Blueprint start-up success calculator	Aplicación creada por la empresa australiana Think Blueprint para calcular la tasa de éxito de una Startup	Yankov, B.

Fuente: Elaboración propia

3.4.5. Discusión

Krishna (2015) y Kakadiya (2016), en sus respectivos trabajos, han identificado factores que influyen en el éxito de las Startups, así, por ejemplo, el primero considera un grupo de factores, tales como Fecha de inicio, tiempo de fundación de la empresa, capital semilla, entre otros, que son importantes en el proceso de crecimiento de una Startup. Por su lado, Kakdiya, S. hace una categorización de los factores en tres tipos como atributos del equipo, de los miembros del equipo y del entorno, mas no incluye factores relacionados a los recursos económicos.

Con relación a los modelos de predicción, Krishna (2015) plantea un modelo basado en dos componentes: preprocesamiento y procesamiento.

Si bien es cierto que las condiciones de cada país tienen sus particularidades, es importante incluir otras variables que no están en los estudios, por lo que existe algunas limitaciones en los trabajos analizados. Por otro lado, sabiendo que en el ciclo de vida de una Startup de TI existen fases, también es necesario analizarla desde este punto de vista.

CAPÍTULO 3: MÉTODO Y MODELO PREDICTIVO DEL ÉXITO DE LAS STARTUPS DE TI

En este capítulo, se presenta un método y un modelo predictivo del éxito de las Startups de TI. El método costa de 4 procesos definidos y el modelo implementa el método a través de 7 técnicas individuales de machine learning. Este apartado está organizado en tres secciones: (1) Método para generar modelos predictivos del éxito de Startups de TI, (2) Modelo predictivo del éxito de Startups de TI y (4) Implementación del modelo predictivo a través de un sistema web.

3.1. Método para generar modelos predictivos del éxito de Startups de TI

Se propone un método sistemático para generar un modelo de ML de pronóstico del éxito de un STI, que considera 4 procesos secuenciales relacionados: selección de factores, extracción de datos, preprocesamiento, y aprendizaje. El método inicia con el proceso de selección de factores y, a partir de ello, se extraen los datos desde diversas fuentes, seguidamente, se preprocesan los datos, y con ello se realiza el proceso de aprendizaje que tiene como salida el modelo pronóstico de ML (ver Figura 14).

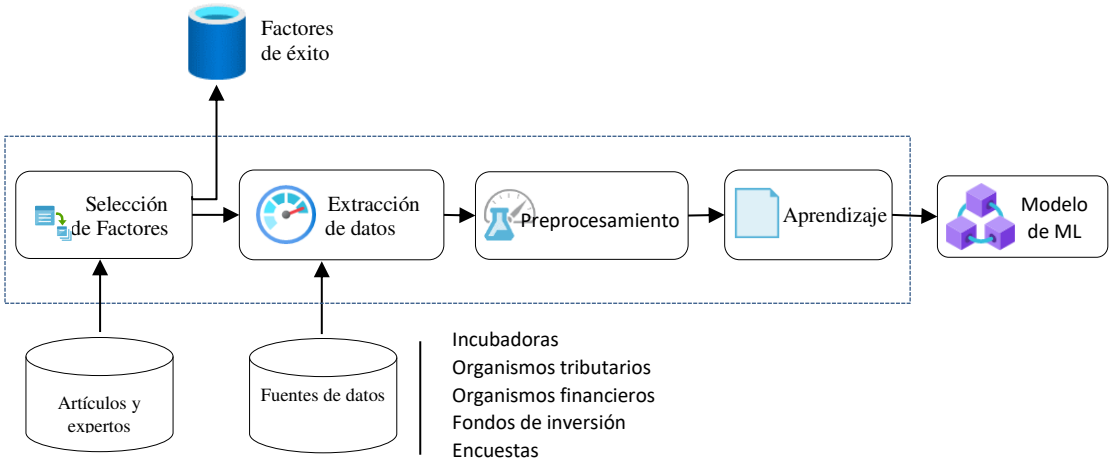

Figura 14. Método para generar modelo de Machine Learning de predicción del éxito de una Startup

Fuente: Elaboración propia

3.1.1. Selección de factores

Se seleccionan los factores que influyen en el éxito (ver Apéndice A) considerando el contexto del estudio (pues los factores pueden corresponder a realidades y periodos diferentes, por lo que podrían no ser válidos) y la disponibilidad de datos significativos. Algunas técnicas para la selección de factores son el Análisis de Componentes Principales (PCA) (Shlens, 2014), Fast Fourier Transform (FFT) (Chandrashekar & Sahin, 2014), Algoritmo GreedySptepwise (Pouhaschemi, S. & Mashalizadeh, A; 2013), Algoritmo Forward y Algoritmo Backward (González, A; 2015).

3.1.2. Extracción de datos

Este proceso consiste en la obtención de datos confiables de los factores seleccionados de muchos STIs (con o sin éxito) desde diversas fuentes como incubadoras de empresas, fondos de inversión, programas gubernamentales,

organismos tributarios y encuestas. También se puede usar dataset de la literatura, siempre que estas se aproximen al contexto de estudio.

3.1.3. Preprocesamiento

Este proceso consiste en transformar los datos obtenidos en el proceso de extracción de datos, en datos que se puedan usar por los modelos y algoritmos de ML. Para este fin, existen diversos subprocesos, tales como limpieza de datos (Corrales, et al., 2020), imputación de valores (Useche & Mesa, 2006), categorización (Kampen, 2019), normalización (Singh, 2019), y balanceo de datos (Kamiran & Calders, 2012).

3.1.4. Aprendizaje

Este proceso recibe como entrada los datos preprocesados, y mediante tres subprocesos (entrenamiento-validación, calibración y prueba) genera un modelo de ML de pronóstico del éxito de un STI (ver Figura 15). Los datos preprocesados se particionan en dos: “TV-Data” y “Test-Data”. El primero es usado por el proceso “Entrenamiento-Validación”, en el cual un algoritmo de ML es entrenado con gran parte de estos datos y luego es validado (por ejemplo, la validación cruzada) con los datos restante. Si el resultado obtenido es satisfactorio, se pasa al proceso de “Prueba”, de lo contrario, se realiza el proceso “Calibración” y se repite el proceso. En Prueba, el modelo obtenido en el proceso previo se pone a prueba con los datos de “Test-Data”, si el resultado es satisfactorio, entonces se obtiene el modelo de pronóstico, de lo contrario, se realiza el proceso “Calibración” y se retorna al proceso de “Entrenamiento-Validación”. El proceso de Calibración consiste en el ajuste de los hiper parámetros del algoritmo de ML, proceso que está automatizado en muchas librerías para ML.

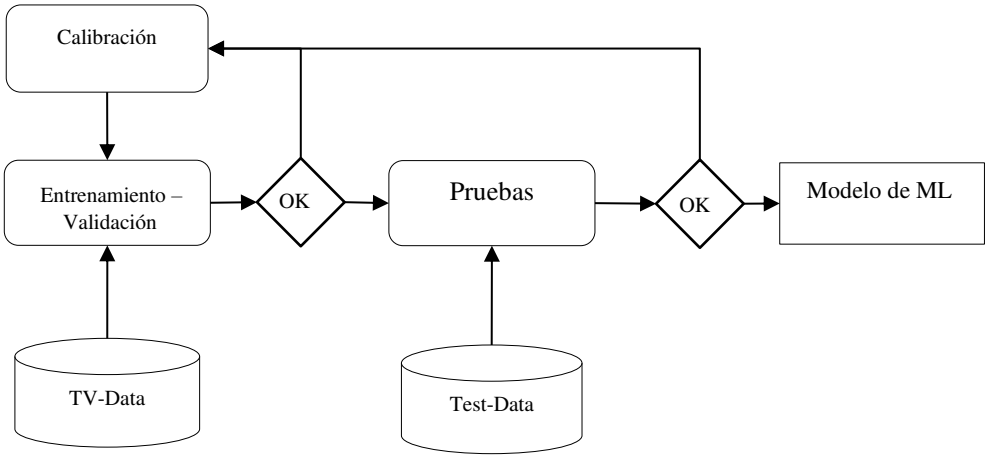

Figura 15. Diagrama del proceso de aprendizaje del modelo de Machine Learning

Fuente: Elaboración propia

3.2. Modelo de predicción del éxito de Startups de TI

Basado en el modelo de ML obtenido por el método propuesto, se propone un modelo para el pronóstico del éxito de una Startup denominado Information Technology Startup Prediction Model (ITSPM) que considera 3 procesos: extracción de datos, preprocesamiento y pronóstico. Los procesos están interrelacionados y

siguen una secuencia de ejecución.

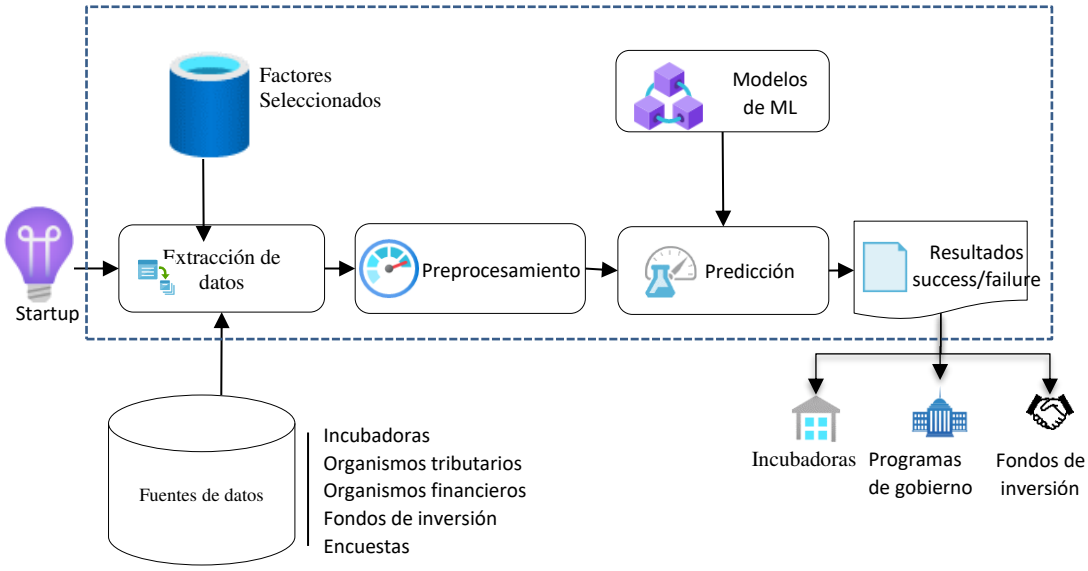

Figura 16. Modelo para la predicción del éxito de una Startup de TI

Fuente: Elaboración propia

El modelo recibe una Startup a ser evaluada, seguidamente, mediante un proceso de Extracción de datos, se extraen, desde diversas fuentes, los datos asociados a los factores de éxito obtenidos por el método que, a su vez, pasan por una fase de Preprocesamiento para obtener datos consistentes y adecuados, los cuales son remitidos al proceso de Pronóstico donde se usa el modelo de Machine Learning obtenida por el método y predice el éxito o fracaso del STI. Este resultado es importante para el análisis y la toma de decisión por parte de los agentes, tales como incubadora, fondo de inversión, inversores, programas gubernamentales de emprendimiento y el emprendedor.

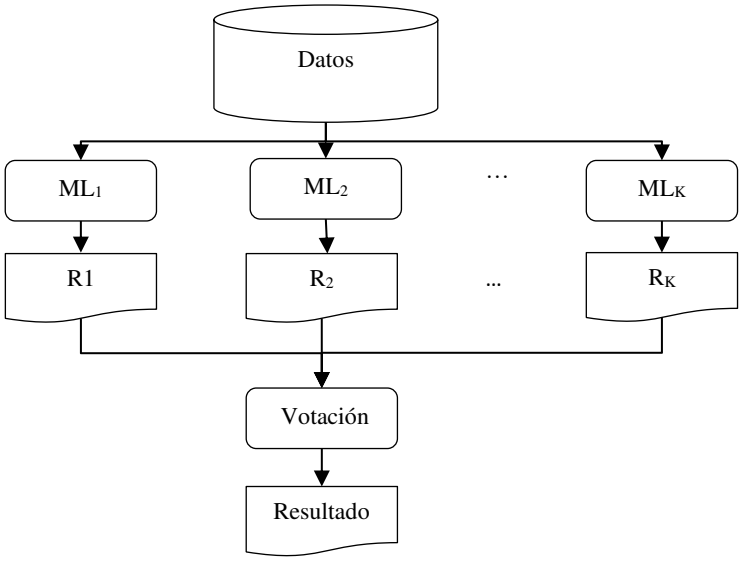

Figura 17. Modelo híbrido de machine learning de predicción del éxito de una Startup

Fuente*:* Elaboración propia

Los procesos de Extracción de datos y Preprocesamiento son los mismos dados en el método con la diferencia que los datos corresponde al STI a ser evaluada. Para el proceso de Pronóstico, se ha considerado un modelo híbrido (ver Figura 17), que consta de un número impar de modelos de ML (ML_1, ML_2, …ML_k) obtenidos por el método al aplicarse "k" veces a diferentes algoritmos de ML, y una estrategia de decisión, como la votación, es decir, el resultado que presenta la mayoría de los modelos. Esta estrategia, en general, presenta mejores resultados que los modelos ML por separado, esto se explica porque la probabilidad de error de un modelo híbrido se reduce a medida que los resultados de la mayoría de los modelos de ML que la componen coinciden.

3.3. Implementación del modelo

Para la implementación del modelo de predicción ITSPM a partir del método propuesto, se ha desarrollado un software web, como herramienta práctica para predecir el éxito de una Startup de TI, en función a los factores críticos de éxito seleccionados de la literatura y de un dataset en particular. Para el desarrollo del sistema, se ha tomado en cuenta las fases de desarrollo de software, tales como elicitación de requisitos, diseño detallado, diseño de arquitectura e implementación basada en una adaptación propia de metodologías agiles.

3.3.1. Elicitación de requisitos

Requisitos funcionales

Los requisitos funcionales son un conjunto de atributos que describen el comportamiento del sistema, considerando los actores que intervienen en cada acción, así como también las restricciones, prioridades, entradas y salidas. En la Tabla 7, se muestran los requisitos funcionales del sistema de predicción del éxito de Startups de TI.

Tabla 7. *Requisitos funcionales para el desarrollo del software*

Código	Nombre	Descripción
RF-001	Registrar usuario	El sistema permite registrar usuarios con los datos de correo, usuario, nombres y contraseña.
RF-002	Registrar Startup	Se registra la información de una Startup con los siguientes campos: nombre de la Startup, edad, sector industrial, localización, así como los

		datos del fundador: Nombre, Apellidos, Género y Formación.
RF-003	Registrar factores críticos del éxito	El sistema permite el registro de los Factores Críticos de Éxito seleccionados de diferentes fuentes de datos y priorizados para un dominio específico.
RF-004	Seleccionar modelo de predicción	Se selecciona la Startup a evaluar previamente registrada, así como el modelo de ML previamente entrenado para la predicción.
RF-005	Generar predicción del éxito	Se realiza el proceso de predicción con los datos registrados de Factores, Startup y Modelo de ML.

Fuente: Elaboración propia

Requisitos no funcionales

Los requisitos no funcionales también denominados atributos de calidad, son considerados elementos condicionantes del comportamiento (opcionales o deseables) de un software que permiten satisfacer las necesidades de los usuarios, además de ello, son el input para el diseño de la arquitectura. En la Tabla 8, se muestra estos requisitos no funcionales del sistema de predicción del éxito de Startups de TI.

Tabla 8. *Requisitos no funcionales para el desarrollo del software*

Código	Nombre	Descripción	Prioridad
RNF-001	Usabilidad	El sistema debe ser atractivo para el usuario y contar con una documentación clara y precisa.	Alta
RNF-002	Confiabilidad	El sistema debe ser tolerante a fallos y capacidad de recuperación ante los mismos.	Alta
RNF-003	Performance	El sistema tendrá un tiempo de respuesta < a 10 segundos en cada petición.	Alta
RNF-004	Mantenibilidad	El sistema debe entendible y fácil de modificar.	Media
RNF-005	Escalabilidad	El sistema debe permitir la escalabilidad horizontal a fin de responder ante los cambios futuros: 10 % por cada año.	Media
RNF-005	Seguridad	El sistema deberá contener seguridad a nivel de autenticación de usuarios a través usuario y contraseña protegido con Json Web Token (JWT) u oAuth 2.0.	Alta
RNF-005	Disponibilidad	El sistema debe estar en funcionamiento 24x7 y garantizar una disponibilidad del 99.9 %.	Media

Fuente: Elaboración propia

3.3.2. Casos de uso

Los casos de uso del sistema son los siguientes: Registrar Usuarios, donde el propietario del caso es el administrador del sistema, el mismo que podrá hacer el registro, actualización y borrado de los usuarios de la plataforma; Registrar Startups, que incluye registrar factores y hacer las predicciones para cada Startup; finalmente, se tiene el caso Generar Reportes. La Figura 18 muestra los principales casos de uso del sistema.

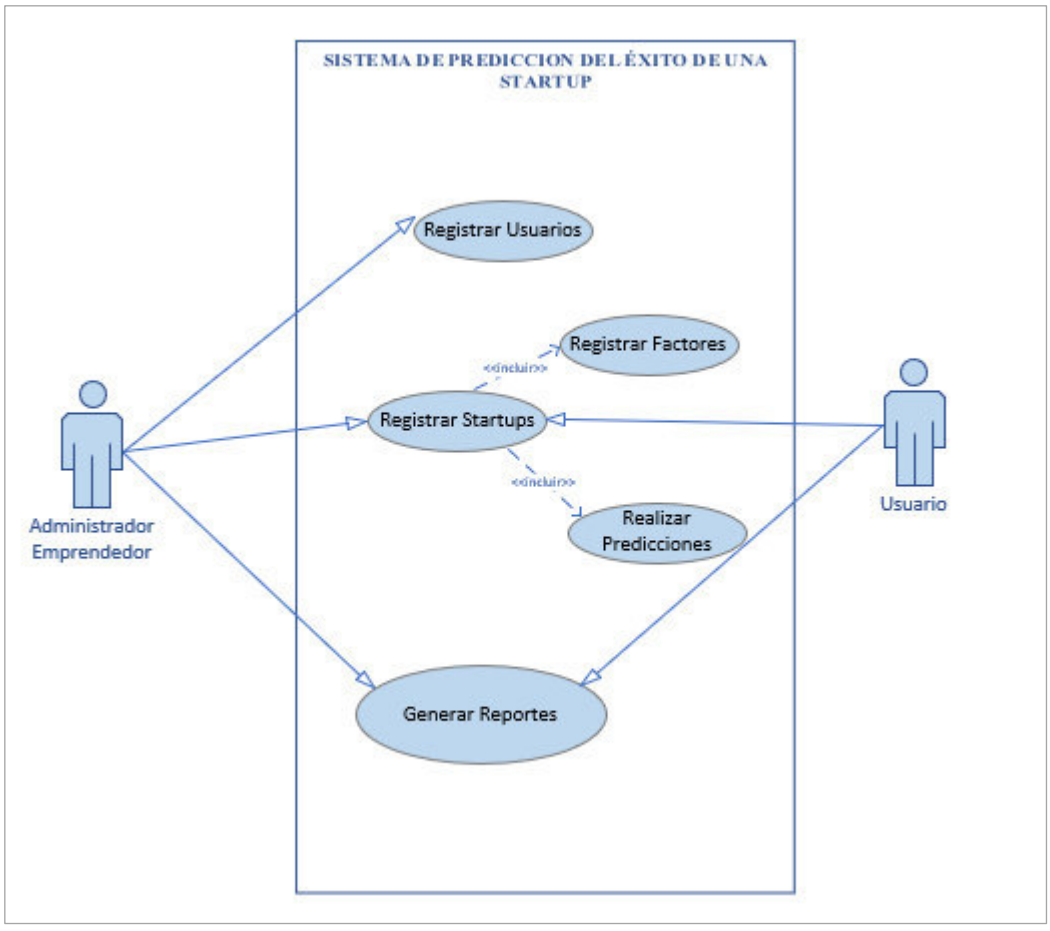

Figura 18. Diagrama de casos de uso del sistema

Fuente: Elaboración propia

3.3.3. Diseño

Diagrama de clases

El diseño comprende la representación de las entidades, atributos y métodos, así como también los diagramas de secuencia que representa el flujo de la información y mensajes de los diferentes actores y artefactos del sistema. En la Figura 19, se muestra el diagrama de clases del sistema.

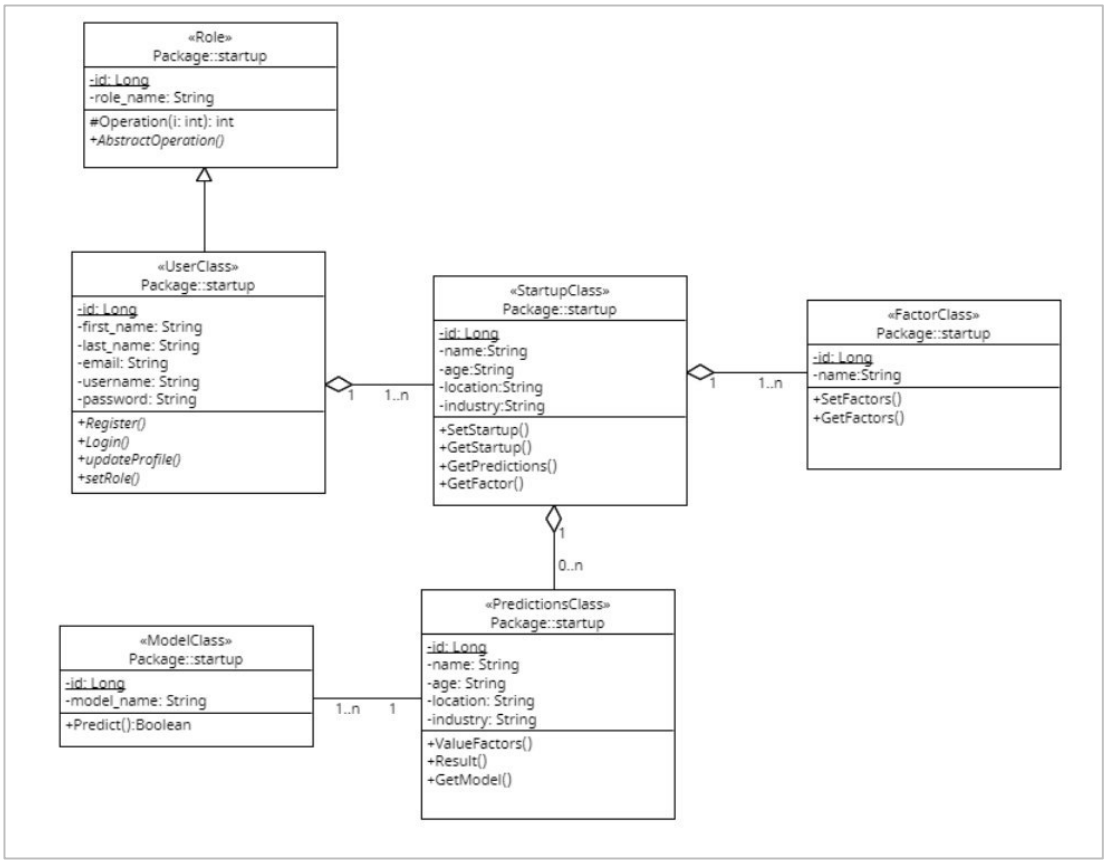

Figura 19. Diagrama de clases del sistema

Fuente: Elaboración propia

Modelo de datos

Entidad Usuario

Entidad que contiene los atributos de los usuarios del sistema, estos pueden ser Administrador, Emprendedor o Responsable de hacer la predicción. En la Tabla 9, se muestra el detalle de la entidad Usuario.

Tabla 9. *Atributos de la Entidad Usuarios*

Atributo	Tipo	Descripción
user_id	Entero (Int)	Identificador único del usuario de tipo auto incremental
first_name	Cadena (varchar)	Nombres del usuario
last_name	Cadena (varchar)	Apellidos del usuario
email	Cadena (varchar)	Email del usuario
password	Cadena (varchar)	Contraseña del usuario de tipo alfanumérico
created_at	Fecha y Hora (timestamp)	Campo de auditoría que registra la fecha y hora de creación de un usuario
updated_at	Fecha y Hora (timestamp)	Campo de auditoría que registra la fecha y hora de actualización de un usuario

Fuente: Elaboración propia

Entidad Rol

Entidad que contiene los atributos del tipo de usuarios del sistema, estos pueden ser Admin, User o Invited. En la Tabla 10, se muestra el detalle de la entidad.

Tabla 10. *Atributos de la Entidad Rol*

Atributo	Tipo	Descripción
role_id	Entero (int)	Identificador único del registro de tipo auto incremental
role_name	Cadena (varchar)	Nombre del Rol que puede ser Admin, Usuario o Invited

Fuente: Elaboración propia

Entidad Startups

Entidad que contiene los atributos de las Startups de TI que van a ser sometidas al modelo de predicción. En la Tabla 11, se muestra el detalle de la entidad.

Tabla 11. *Atributos de la Entidad Startup*

Atributo	Tipo	Descripción
startup_id	Entero (int)	Identificador único del registro de tipo auto incremental
startup_name	Cadena (varchar)	Nombre de la Startup que va a ser evaluada
age	Entero (int)	Edad de la Startup en años
location	Cadena (varchar)	Ubicación geográfica de la Startup
industry	Cadena (varchar)	Sector industrial al que pertenece la Startup

Fuente: Elaboración propia

Entidad Predicciones

Entidad que contiene los atributos de las predicciones de cada Startup de TI que va a ser sometidas al modelo de predicción. En la Tabla 12, se muestra el detalle de la entidad.

Tabla 12. *Atributos de la Entidad Predicciones*

Atributo	Tipo	Descripción
prediction_id	Entero (int)	Identificador único del registro de tipo auto incremental
startup_id	Entero (int)	Clave foránea que relaciona con la Entidad Startup
model_id	Entero (int)	Clave foránea que relaciona con la Entidad Modelo
factor_value	Entero (int)	Valor del factor asociado a la Startup evaluada
result_prediction	Entero (int)	Resultado de la predicción que puede ser 0= Failure, 1= Success

Fuente: Elaboración propia

Entidad Factores

Entidad que contiene los atributos de los factores críticos de éxito seleccionados para la predicción. En la Tabla 13, se muestra el detalle de la entidad.

Tabla 13. *Atributos de la Entidad Factores*

Atributo	Tipo	Descripción
factor_id	Entero (int)	Identificador único del registro de tipo auto incremental

factor_name	Cadena (int)	Nombre de los factores críticos de éxito seleccionados

Fuente: Elaboración propia

Entidad Modelo

Entidad que contiene los atributos de los factores críticos de éxito seleccionados para la predicción. En la Tabla 14, se muestra el detalle de la entidad.

Tabla 14. *Atributos de la Entidad Modelo*

Atributo	Tipo	Descripción
model_id	Entero (int)	Identificador único del registro de tipo auto incremental
model_name	Cadena (int)	Nombre del modelo de predicción entrenado y que es usado para la predicción de una Startup.

Fuente: Elaboración propia

La Figura 20 muestra el Diagrama de Entidad –relación del sistema de predicción del éxito de las Startups de TI– con sus respectivos atributos y relaciones.

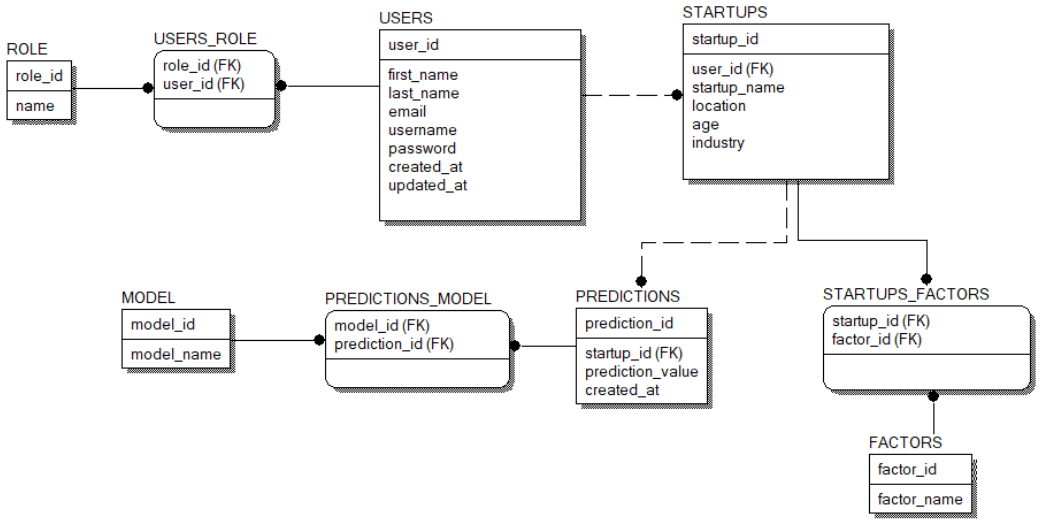

Figura 20. Diagrama de Entidad-Relación del sistema

Fuente: Elaboración propia

Interfaces de usuario

Las interfaces de usuario a nivel de diseño muestran la presentación visual del comportamiento del sistema en la primear iteración, las mismas que son implementadas en el front-end del sistema. Las principales interfaces se muestran en desde la Figura 21 a la Figura 27.

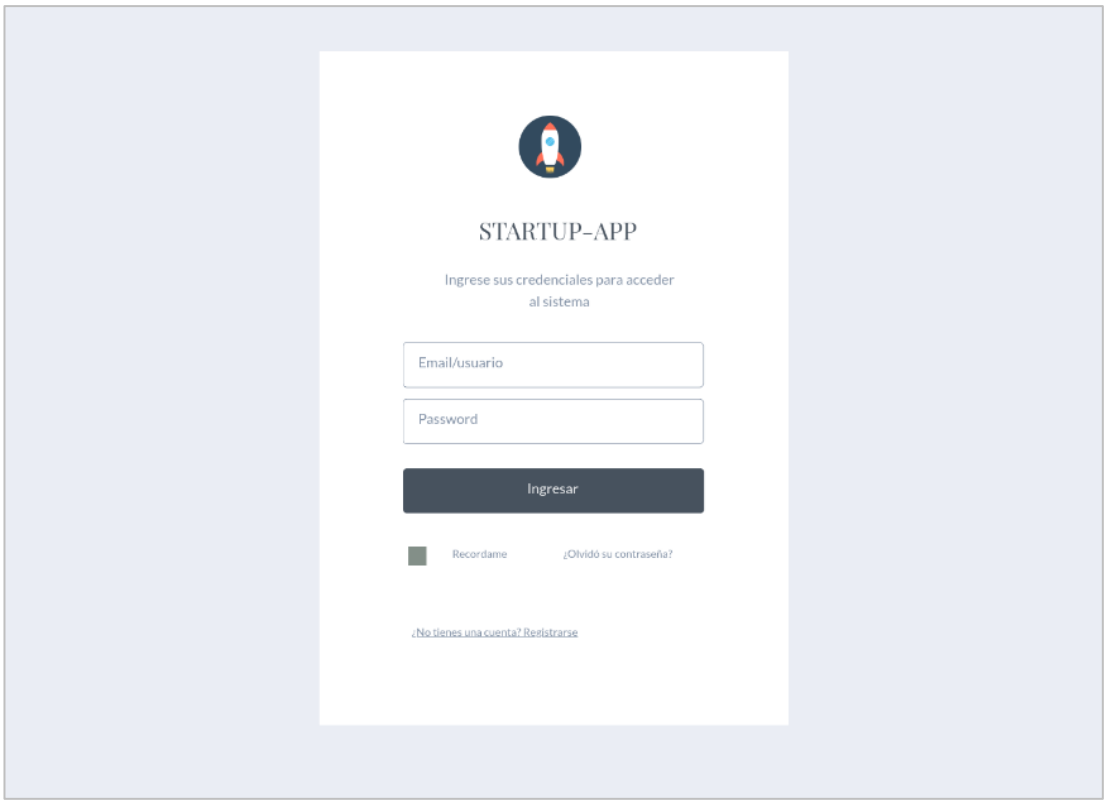

Figura 21. Wireframe de interfaz de usuario de Login

Fuente: Elaboración propia

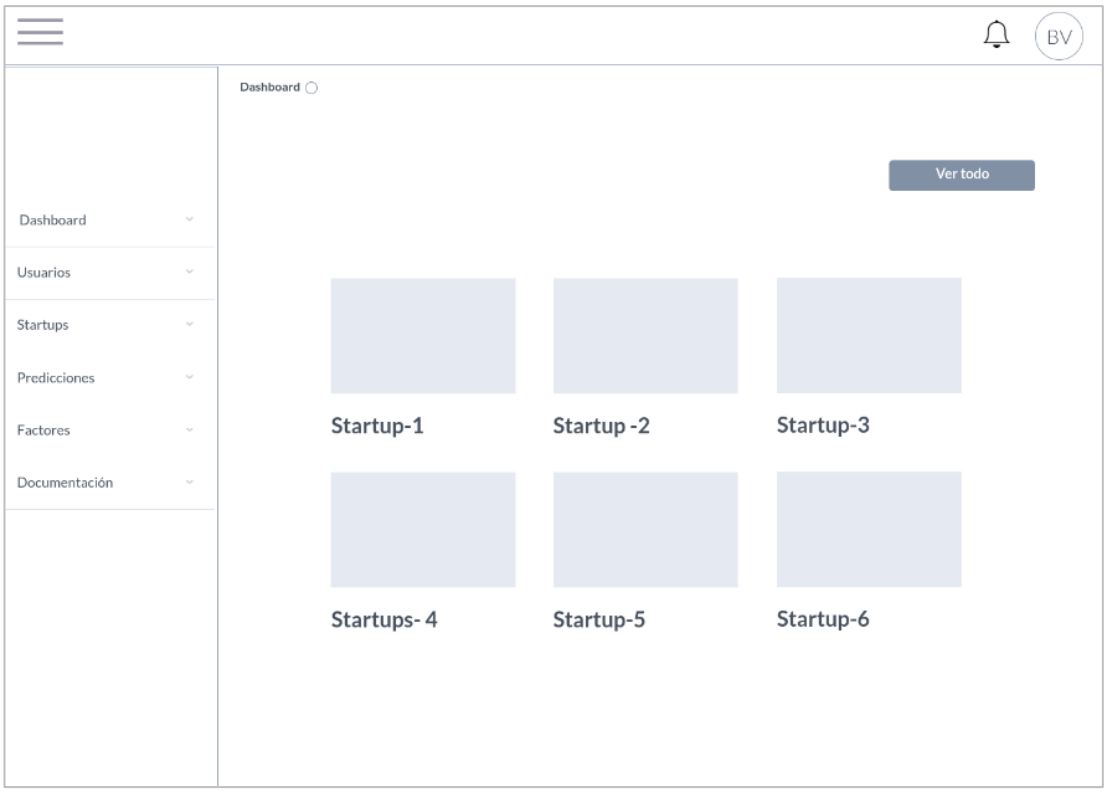

Figura 22. Wireframe de interfaz de usuario de dashboard

Fuente: Elaboración propia

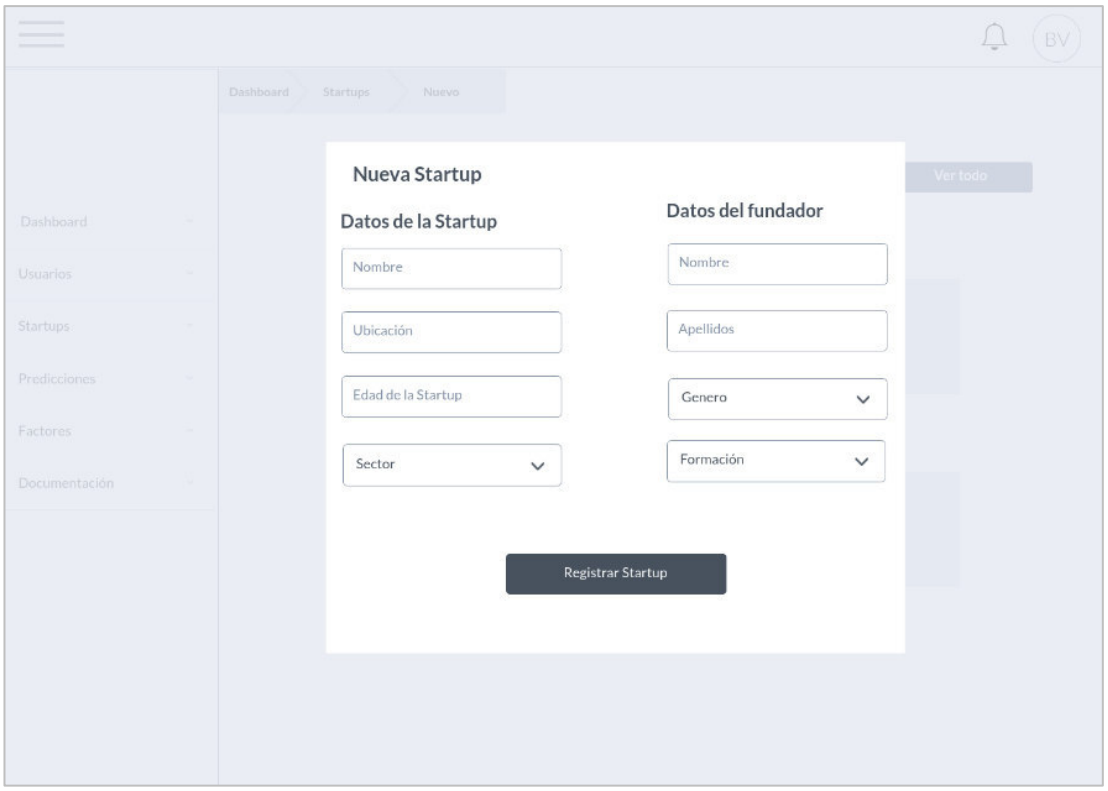

Figura 23. Wireframe de interfaz de usuario Registro de Startup

Fuente: Elaboración propia

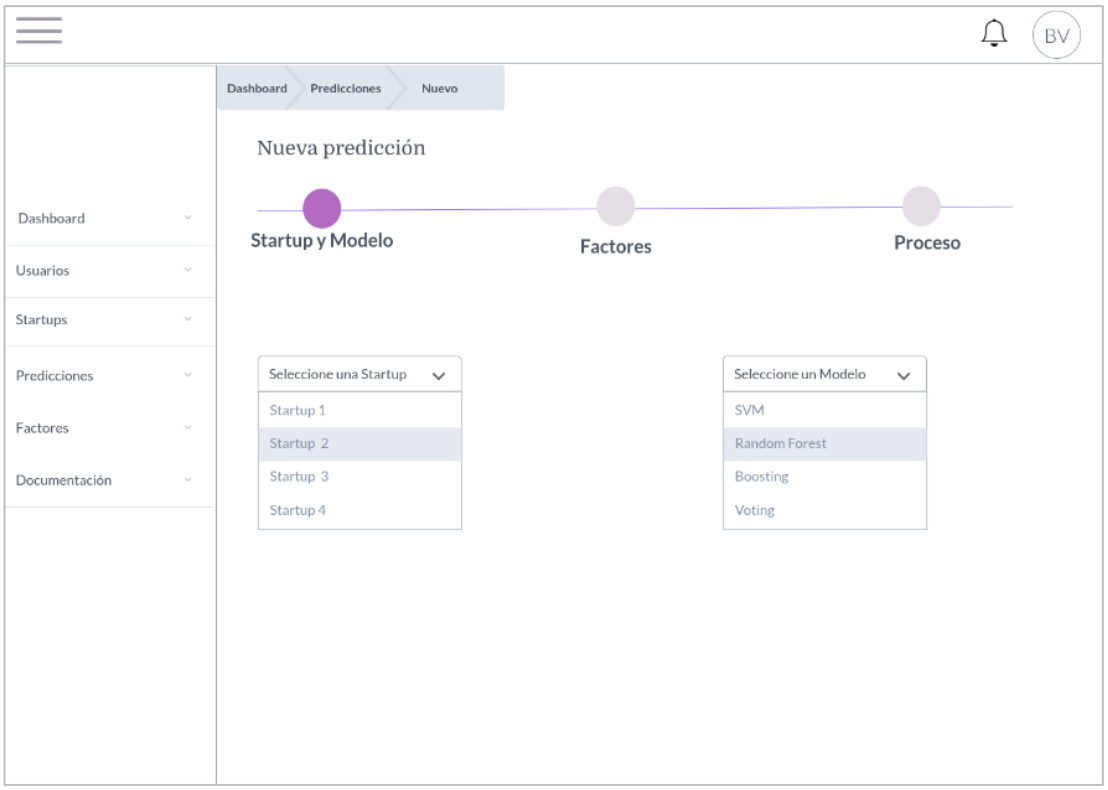

Figura 24. Wireframe de interfaz de usuario de Predicción (selección de Startups y modelo)

Fuente: Elaboración propia

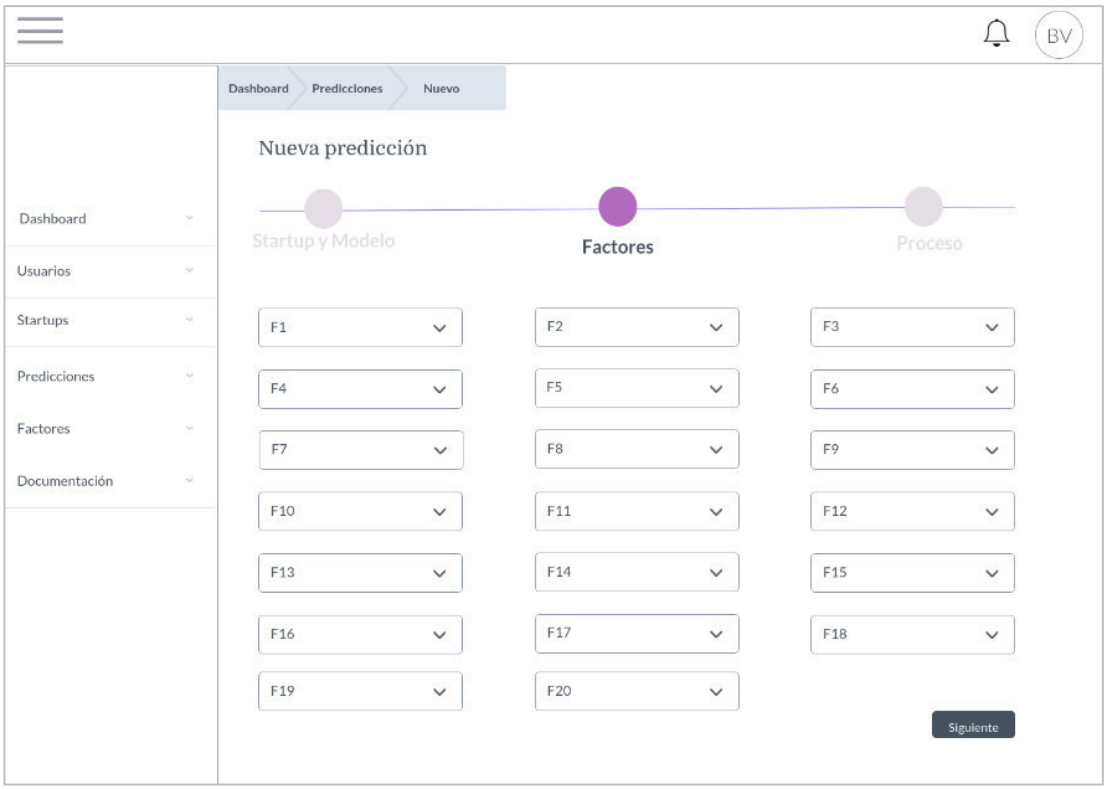

Figura 25. Wireframe de interfaz de usuario de Predicción (Registro de factores)

Fuente: Elaboración propia

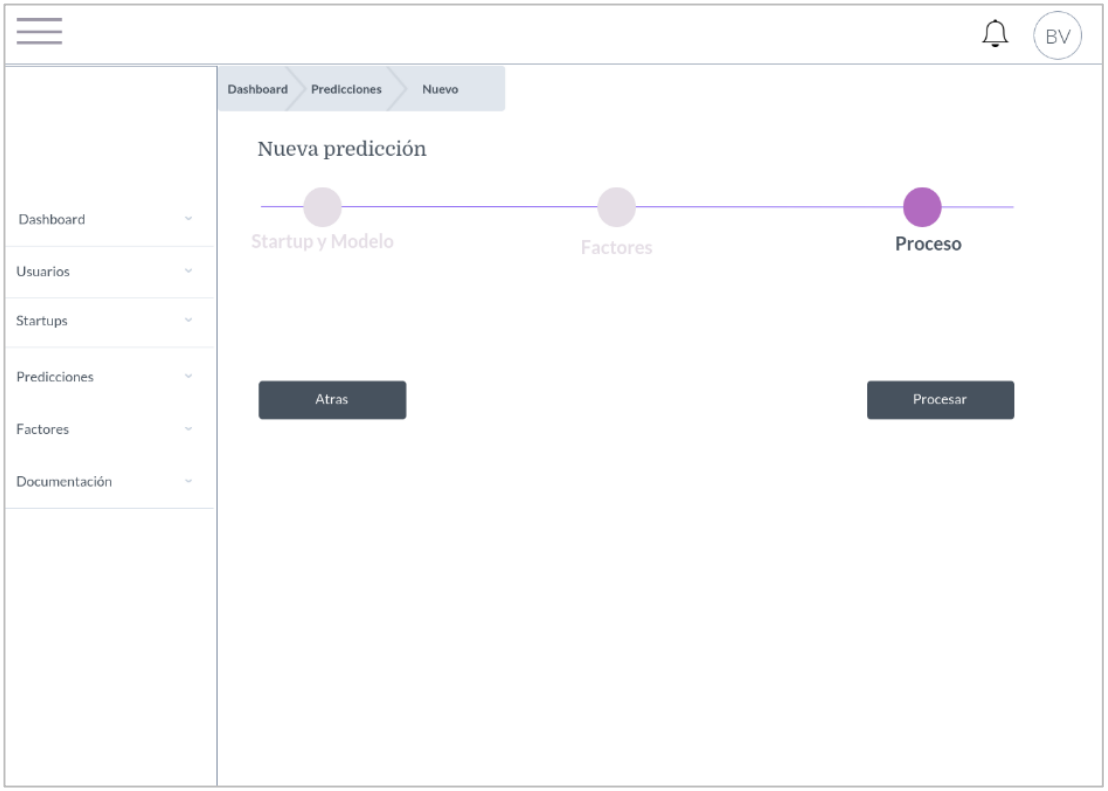

Figura 26. Wireframe de interfaz de usuario de Predicción (Procesamiento)

Fuente: Elaboración propia

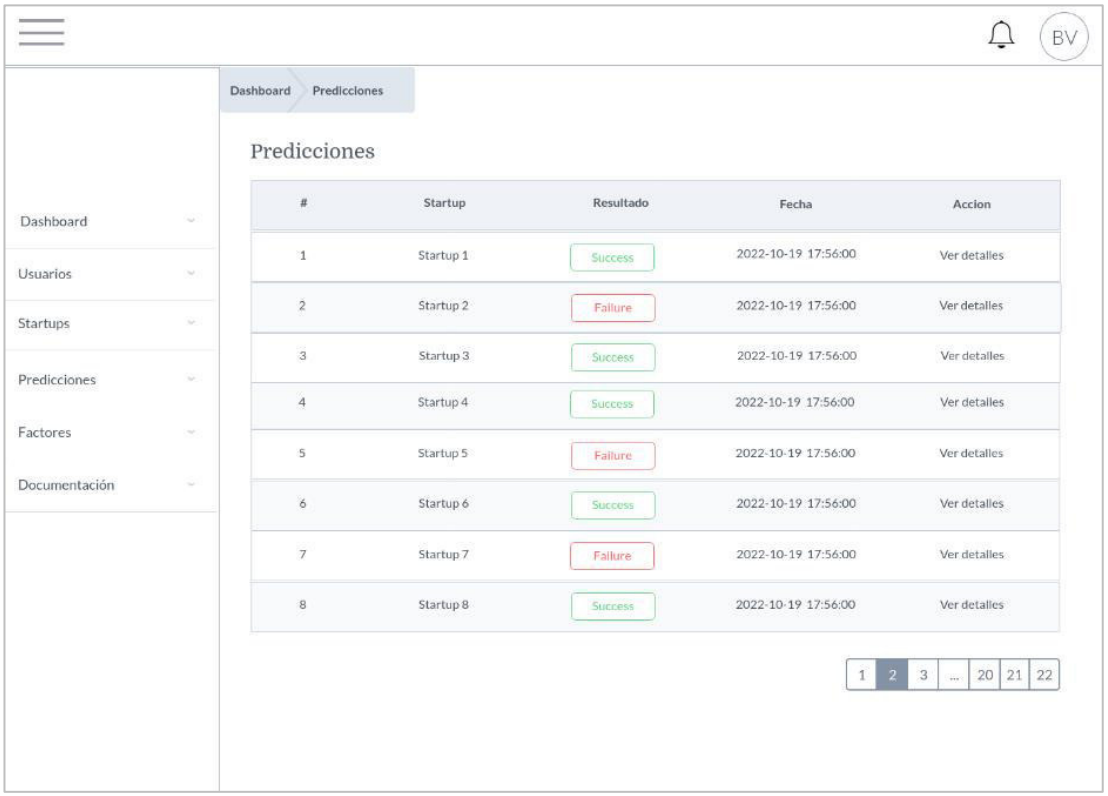

Figura 27. Wireframe de interfaz de usuario de Resultados de Predicción

Fuente: Elaboración propia

3.3.4. Arquitectura

Para el diseño de la arquitectura del sistema de predicción del éxito de startups de TI, se ha tomado en cuenta las restricciones y concerns dados en los requisitos funcionales y no funcionales, con la finalidad de plantear un diseño arquitectura acorde al dominio del tema. Además, consideramos como punto de partida el modelo de vistas de arquitectura de software 4+1 propuesto por Kruchten, P (1995). En la Figura 28, se muestra la representación arquitectural del sistema.

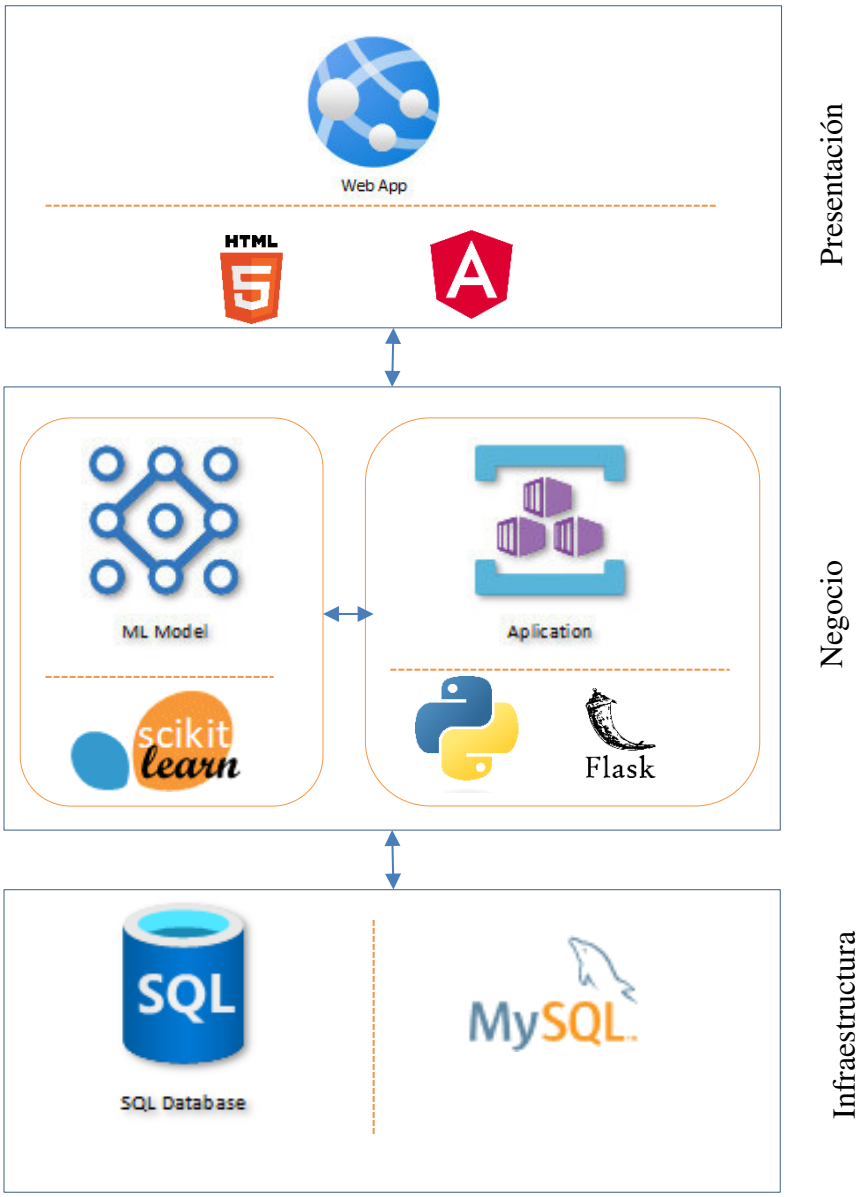

Figura 28. Arquitectura de referencia propuesta del sistema

Fuente: Elaboración propia

En cuanto al despliegue del sistema, este se realiza tomando en cuenta lo siguiente: Servidor de Web, que es el encargado de alojar los archivos del front-end y que a su vez hace uso de CDN y archivos de configuración propios; Servidor de

Aplicaciones 1, que comprende los modelos de Machine Learning previamente entrenados, nombrados y versionados en formato .pkl; Servidor de aplicaciones 2, que contiene los componentes lógicos de la aplicación; y Servidor de base de datos, que incluye el modelo relacional de las tablas en lenguaje SQL. En la Figura 29, se muestra el diagrama de despliegue del sistema.

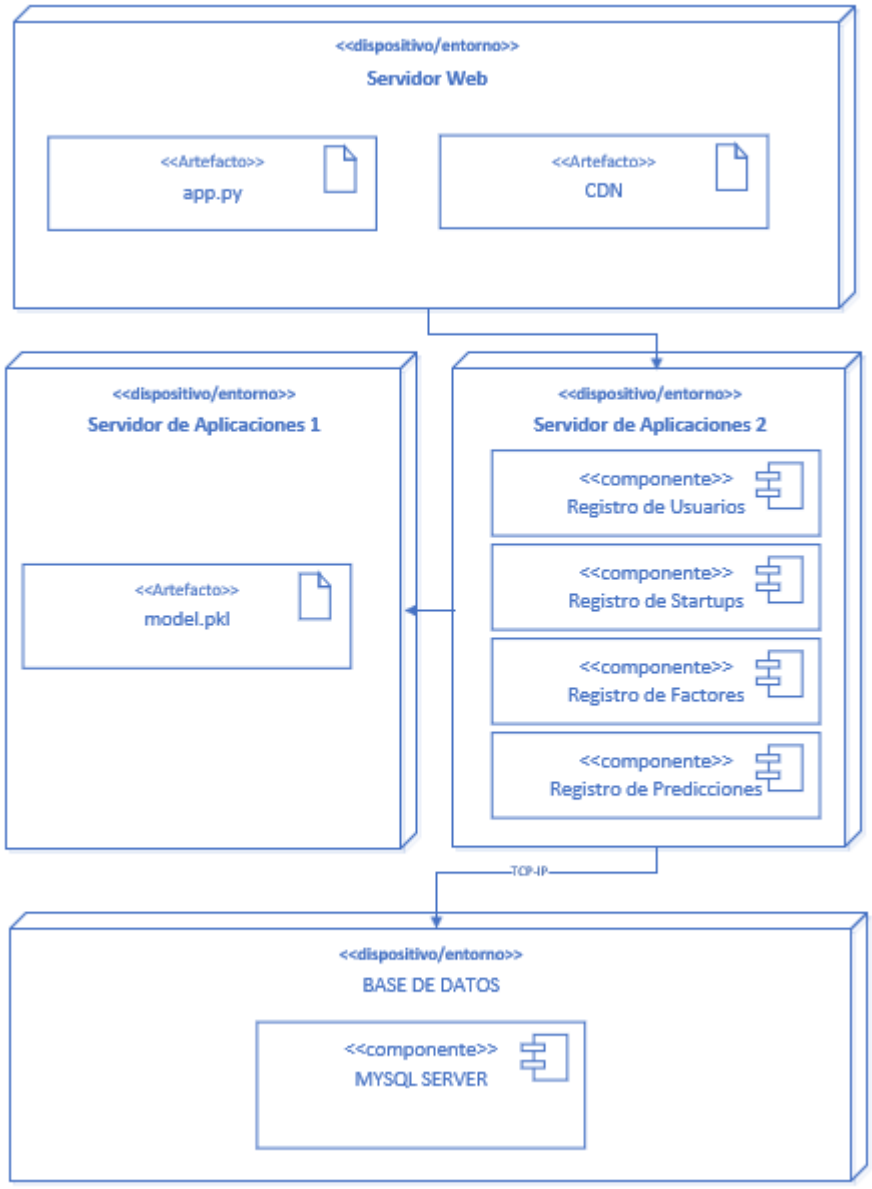

Figura 29. Diagrama de despliegue del sistema

Fuente: Elaboración propia

3.3.5. Funcionamiento

Para las pruebas de funcionamiento del sistema, el usuario previamente registrado, accede a la aplicación a través de sus credenciales de username y password. Una vez dentro del sistema, en primer lugar, se registra un nuevo startup con los campos definidos tal como se muestra en la Figura 30.

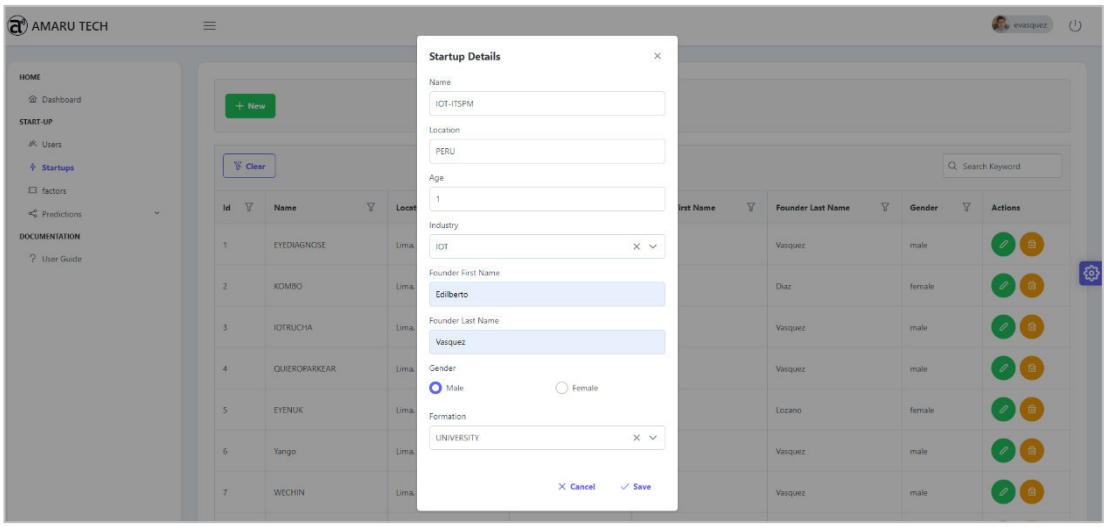

Figura 30. Vista del proceso de registro de un nuevo startup

Fuente: Elaboración propia

Luego de registrar el startup, en el menú de “Predicciones” se selecciona “New”, para desplegar un nuevo formulario donde se selecciona el modelo de ML, el startup previamente registrado y los valores de cada uno de los factores asociados al Startup a predecir, para finalmente hacer el proceso de predicción a través del botón “Process an Save”. Este procedimiento se muestra en la Figura 31.

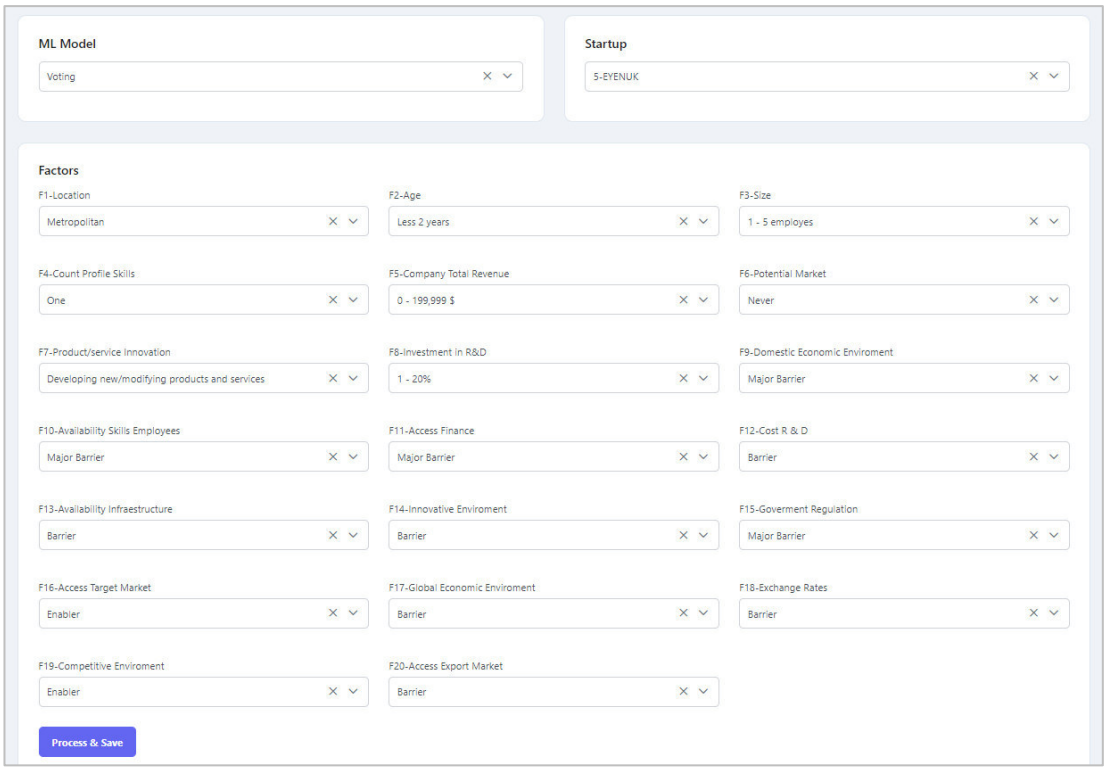

Figura 31. Vista del proceso de selección de Modelo, Startup, Factores y Predicción

Fuente: Elaboración propia

Finalmente, los resultados de las predicciones de muestran en una tabla con los datos de Id, Nombre de la Startups, resultado de la predicción y fecha de la última predicción como en la Figura 32.

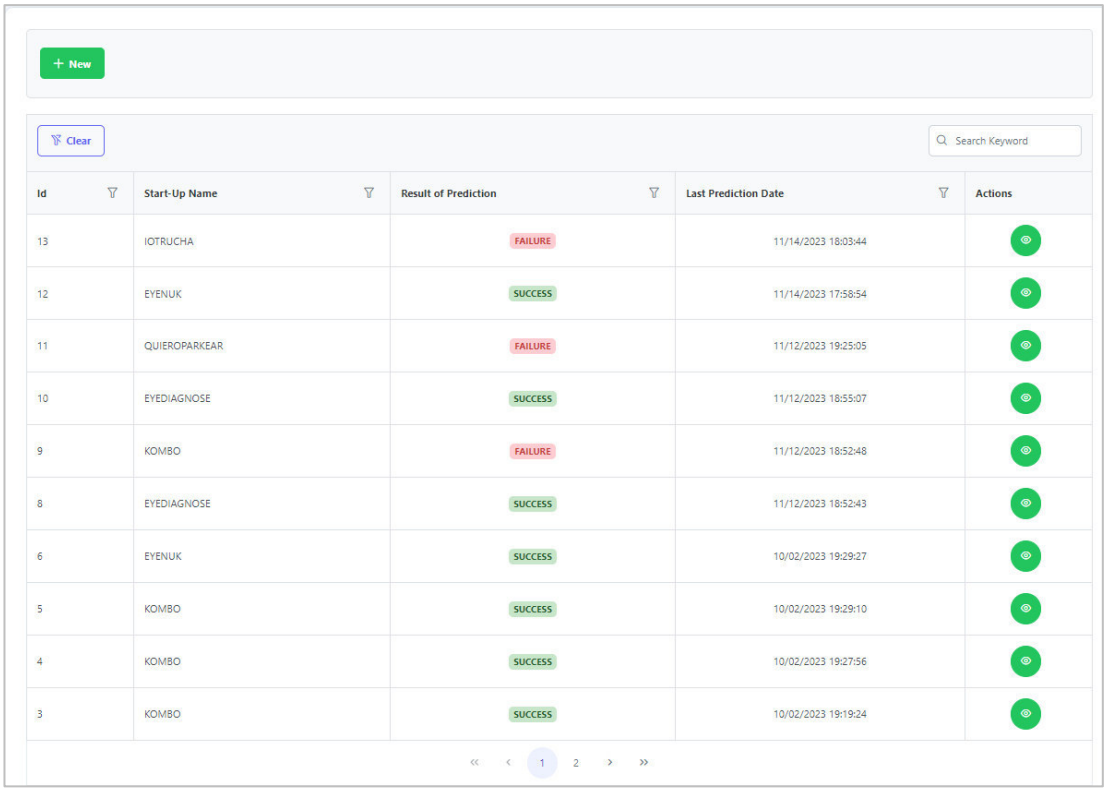

Figura 32. Vista de los resultados de predicción por cada startup.

Fuente: Elaboración propia

CAPÍTULO 4: VALIDACIÓN

Para validar los modelos propuestos, se realizará experimentos numéricos considerando dataset de la literatura, métricas, preprocesamiento como limpieza y balanceo de datos, así como selección de factores, finalmente, se analizará los resultados.

4.1. El dataset

Se considera el dataset usado por Tomy & Pardede (2018) con información del estudio "2013 Victorian ICT Industry Statistics survey" actualizado al 2015 bajo licencia de Creative Commons Attribution 4.0 International, el mismo que consta de 265 registros de igual número de empresas.

Cada registro cuenta con 63 datos sobre 23 factores de la literatura y otros atributos, etiquetados de la siguiente manera: 182 como éxito, 80 como fracaso y 3 no tiene información. En el dataset, destacan las empresas de Software development and installation, systems analysis and computer programming y computer software consulting, con el 53 % del total, donde el éxito es considerado como ser rentable.

4.2. Métricas

Para la evaluación de los resultados de los experimentos numéricos, se definen las métricas de Exactitud, Precisión y Especificidad, cuya formulación se muestra en la Tabla 15.

Tabla 15. *Métricas usadas en el proceso de predicción de las Startups de TI*

Métrica	Descripción	Formulación

Exactitud (ACU)	Tasa de predicciones correctamente clasificadas	$\frac{TP + TN}{TP + FP + TN + FN}$
Precisión (PRE)	Tasa de positivos correctamente clasificados	$\frac{TP}{TP + FP}$
Especificidad (ESP)	Tasa de negativos correctamente clasificados	$\frac{TN}{TN + FP}$

Fuente: Elaboración propia

Donde:

$TP = True\ Positives$, la Startup es exitosa y el algoritmo predice como success;

$TN = True\ Negatives$, la Startup no es exitosa y el algoritmo predice como failure;

$FP = False\ Positives$, la Startup no es exitosa y el algoritmo predice success;

$FN = False\ Negatives$, la Startup es exitosa y el algoritmo predice como failure.

4.3. Preprocesamiento

Con la finalidad de tener datos consistentes y que los resultados finales sean fiables, se han realizado las siguientes actividades de preprocesamiento:

Limpieza de datos

En la limpieza de datos, se identificó 27 columnas con más del 50 % de datos en blanco, por lo que fueron descartas, además, se descartó 16 columnas por similitud de atributo, es decir, se redujo el número de columnas a 20, donde cada una de ellas corresponde a un factor. Además, se identificó 3 registros sin etiqueta de éxito o fracaso y 14 con más del 50 % de datos vacíos, por lo que fueron eliminados, quedando 248 registros válidos.

Tabla 16. *Factores y clases de atributos seleccionados*

ID	*Factor*	*Class*	*ID*	*Factor*	*Class*
F1	Location	A	F11	Financial capital	I
F2	Age	B	F12	R&D	I
F3	Startup size	C	F13	Availability of infrastructure	I
F4	Amount employee skills	D	F14	Innovation environment	I
F5	Company revenue	E	F15	Government regulation	I
F6	Export produtcs	F	F16	Access to target market	I
F7	Innovation of product / service	G	F17	Global economic environment	I
F8	Size of investment	H	F18	Exchange rates	I
F9	Environment	I	F19	Competition	I
F10	Availability skilled employees	I	F20	Access to export market	I

Fuente: Elaboración propia

Clases y porcentaje:

A. *1 = Metropolitan (77%), 2 = Regional (10%), 3 = Interstate (9%), 4 = Overseas (4%)*
B. *1 = Less than 2 years (11%), 2 = 2–4 years/(10%=, 3 = 5–9 years(28%), 4 = 10–19 years(30%), 5 = 20+ years(22%)*
C. *1 = 1–5(34%); 2 = 6-20(34%); 3 = 21–100(21%); 4 = 101–1000(7%); 5 = 1000 + (employees)(4%)*
D. *1(61%),2(20%),3(10%),4(5%),5(5%) (number of skills)*
E. *1 = 0–199,999$ (24%), 2 = 200,000–499,999$(13%), 3 = 500,000–999,999$ (11%), 4 = 1,000,000–4,999,999$(28%), 5 = 5,000,000$+(23%)*
F. *1 = Never (32%, 2 = Irregularly (33%), 3 = Regularly (32%)*
G. *1 = Developing new products and services (80%), 2 = Implementing new or significantly improved operational processes/services (15%), 3 = Implementing new or significantly improved marketing method (3%), 4 = Company not involved in innovation (2%)*
H. *1 = 0–10% (59%), 2 = 11–25% (21%), 3 = 26–50%(4%), 4 = 51–75%(6%), 5 = 75%+(10%)*

I. *1 = Major Enabler (17%), 2 = Enabler (37%), 3 = Neither a barrier nor an enabler (29%), 4 = Barrier (12%), 5 = Major Barrier(5%)*

Balanceo de datos

En 10 registros, se ha completado el dato con el valor entero promedio de su correspondiente factor (imputación de valores). Luego se han balanceado los datos usando Oversammpling, con la finalidad de obtener un mismo número de registros para éxito y fracaso, quedando el dataset con 342 registros (ver Tabla 17), del cual el 10 % (34 registros) se separa para *Prueba* y 90 % (308 registros) se usa para *Entrenamiento-Validación*.

Tabla 17. *Características del dataset original y preprocesado*

Dataset ICT - Australia	Dataset procesado
Número de registros = 265	Número de registros = 342
Número de factores: 23	Número de factores: 20
Clase de éxito = profitabiliy	Clase de éxito = profitabiliy
Tipo de datos: Nominal y numérico	Tipo de datos: Numérico
Número de STI con éxito =182	Número de STI con éxito =171
Número de STI sin éxito = 80	Número de STI sin éxito 171

Fuente: Elaboración propia

4.4. Selección de factores

La selección de factores de un conjunto puede servir para casos en los que se tiene poca información de un conjunto de Startups, así como por temas puntuales de rendimiento. En este caso, se analizan dos técnicas de selección de factores (variables) que son Análisis de Componentes Principales (PCA) y GredyStepWise.

4.4.1. Análisis de componentes principales

La selección de factores se hizo a través de la técnica de Análisis de Componentes Principales (PCA), para ello se usó la rutina implementada en la herramienta WEKA, dando como resultados la matriz de correlación (ver Anexo 1), los autovalores (ver Anexo 2) y auto vectores, así como un ranking de las principales variables a considerar y el porcentaje de la cobertura. En la Tabla 18, se muestra los resultados del proceso.

Tabla 18. *Ranking de atributos generados a través de PCA*

N.° orden	% cobertura	Nuevo factor
1	0.8278	-0.36SIZE_STARTUP- 0.335COMPANY_TOTAL_REVENUE- 0.319ACCESS_TO_FINANCE- 0.3ACCESS_TO_TARGET_MARKET+0.3 INVESMENT_R&D
2	0.7101	0.391COMPANY_TOTAL_REVENUE- 0.389DOMESTIC_ECONOMIC_ENVIROMENT- 0.375GLOBAL_ECONONOMIC_ENVIROMENT+0.3 07SIZE_STARTUP- 0.302AVAILABILITY_OF_SKILLED_EMPLOYEES
3	0.624	0.429AVAILABILITY_INFRAESTRUCTURE+0.388A CCESS_TO_TARGET_MARKET+0.368COUNT_PRO FILES_SKILLS+0.361INNOVATIVE_ENVIROMENT -0.345AVAILABILITY_OF_SKILLED_EMPLOYEES
4	0.5478	0.462EXCHANGE_RATES- 0.353GOVERMENT_REGULATION+0.352COUNT_P ROFILES_SKILLS- 0.342AVAILABILITY_OF_SKILLED_EMPLOYEES- 0.254IN0VATIVE_ENVIROMENT
5	0.485	- 0.557EXPORT_PRODUCTS_CTI+0.373AVAILABILI TY_INFRAESTRUCTURE- 0.368COUNT_PROFILES_SKILLS- 0.325COST_OF_R&D- 0.296MAIN_INNOVATION_ACTIVITIES

6	0.4281	-0.48LOCATION+0.45 AGE_STARTUP+0.356ACCES_TO_EXPORT_MARKETS-0.329GOVERMENT_REGULATION+0.237GLOBAL_ECO0MIC_ENVIROMENT
7	0.3739	0.516MAIN_INNOVATION_ACTIVITIES-0.382EXPORT_PRODUCTS_CTI+0.338IN0VATIVE_ENVIROMENT+0.308COST_OF_R&D-0.291INVESMENT_R&D
8	0.3258	-0.533MAIN_INNOVATION_ACTIVITIES-0.426COMPETITIVE_ENVIROMENT-0.397AGE_STARTUP-0.314INVESMENT_R&D-0.272AVAILABILITY_INFRAESTRUCTURE
9	0.2816	0.35 GOVERMENT_REGULATION-0.316EXPORT_PRODUCTS_CTI-0.314SIZE_STARTUP-0.291MAIN_INNOVATION_ACTIVITIES+0.279COMPETITIVE_ENVIROMENT
10	0.2394	0.486COMPETITIVE_ENVIROMENT+0.476AVAILABILITY_INFRAESTRUCTURE-0.297MAIN_INNOVATION_ACTIVITIES-0.288INVESMENT_R&D-0.242GLOBAL_ECONOMIC_ENVIROMENT
11	0.2033	0.555ACCESS_TO_FINANCE-0.4GLOBAL_ECONOMIC_ENVIROMENT+0.298AVAILABILITY_OF_SKILLED_EMPLOYEES+0.271ACCESS_TO_TARGET_MARKET-0.259DOMESTIC_ECONOMIC_ENVIROMENT
12	0.1712	-0.428INVESMENT_R&D-0.427EXCHANGE_RATES+0.378ACCES_TO_EXPORT_MARKETS-0.295COST_OF_R&D-0.288GOVERMENT_REGULATION
13	0.1412	-0.485DOMESTIC_ECONOMIC_ENVIROMENT+0.352AVAILABILITY_INFRAESTRUCTURE-0.347INVESMENT_R&D+0.321AVAILABILITY_OF_SKILLED_EMPLOYEES+0.282EXPORT_PRODUCTS_CTI
14	0.1118	-0.389IN0VATIVE_ENVIROMENT-0.355LOCATION+0.35 ACCESS_TO_TARGET_MARKET-0.329ACCESS_TO_FINANCE-0.32ACCES_TO_EXPORT_MARKETS

15	0.085	-0.524ACCESS_TO_FINANCE+0.396ACCES_TO_EXPORT_MARKETS-0.331GLOBAL_ECO0MIC_ENVIROMENT+0.311INNOVATIVE_ENVIROMENT+0.276AVAILABILITY_OF_SKILLED_EMPLOYEES
16	0.0597	0.595COUNT_PROFILES_SKILLS+0.392AGE_STARTUP+0.356DOMESTIC_ECONOMIC_ENVIROMENT-0.315EXCHANGE_RATES+0.295AVAILABILITY_OF_SKILLED_EMPLOYEES
17	0.0377	0.629COST_OF_R&D-0.373IN0VATIVE_ENVIROMENT-0.339AVAILABILITY_OF_SKILLED_EMPLOYEES+0.302ACCES_TO_EXPORT_MARKETS-0.234EXPORT_PRODUCTS_CTI

Fuente: Elaboración propia

4.4.2. Greedy Step Wise

Es la técnica de selección de factores o variables de un conjunto grande de datos a partir de la selección de óptimos locales en cada paso. Tiene 5 pasos definidos que se mencionan a continuación.

Inicialización: Comienza con un modelo vacío, es decir, sin variables predictoras seleccionadas.

Etapa hacia adelante (Forward Step): En cada iteración, se evalúa el impacto de agregar una variable predictora adicional al modelo actual. Se selecciona la variable que produce la mejora más significativa según un criterio definido, como la reducción del error cuadrático medio (MSE).

Etapa hacia atrás (Backward Step): Una vez que se ha agregado una variable al modelo, se evalúa el impacto de eliminar cada variable individual del modelo. Se elimina la variable que produce la menor disminución en la medida de rendimiento

elegida. Esto se repite iterativamente hasta que la eliminación de cualquier variable empeore significativamente el rendimiento del modelo.

Etapa de actualización (Update Step): Una vez que se ha eliminado una variable del modelo, se evalúa el impacto de agregar nuevamente las variables previamente eliminadas. Si alguna de las variables eliminadas mejora el rendimiento del modelo, se vuelve a agregar.

Criterio de parada: El proceso de selección de variables continúa iterativamente agregando o eliminando variables hasta que se cumpla algún criterio de parada predefinido. Esto podría ser alcanzar un número máximo de variables seleccionadas, una mejora mínima en la medida de rendimiento o cualquier otro criterio establecido por el usuario.

Para el conjunto de factores del dataset seleccionado para la predicción, aplicando la técnica Greedy Step Wise, arroja 5 factores (ver Tabla 19) que tienen gran influencia en la predicción del éxito en las Startups de TI.

Tabla 19. *Factores seleccionados por el método Greedy Step Wise*

N.° orden	Factor seleccionado
F1	SIZE_STARTUP
F2	COMPANY_TOTAL_REVENUE
F3	INVESMENT_R&D
F4	ACCESS_TO_FINANCE
F5	GLOBAL_ECONOMIC_ENVIROMENT

Fuente: Elaboración propia

4.5. Resultados

Los resultados de los experimentos se han obtenido sobre cuatro escenarios: el primero consta de 23 factores considerando un preprocesamiento parcial; en el segundo, se ha considerado 20 factores usando siete algoritmos y 3 modelos híbridos de Machine Learning; en el tercer escenario, se considera 17 factores, obtenidos por el PCA; y, en el cuarto escenario, se ha considerado 5 factores, obtenidos por el algoritmo GreedyStepWise. Las tablas 20, 21 y 22 muestran los resultados de la validación cruzada para el 90 % del total de registros en cada uno de los escenarios.

Tabla 20. *Resultados de la validación cruzada para el escenario de 20 factores con 10-fold*

Modelo	Acu	Pre	Esp
SVM	94.80	90.79	100.0
Perceptron Multilayer	90.89	96.52	84.67
Naive Bayes	71.02	71.20	76.88
Decision Tree	69.02	67.52	78.17
KNN	75.83	75.92	77.04
Random Forest	88.96	96.43	88.80
Gradient Boosting	90.55	95.16	86.33
Voting3	92.82	93.31	92.96
Voting5	92.82	93.47	92.96
Voting7	92.88	94.82	91.08

Fuente. Elaboración propia

De la Tabla 20, se observa que los mejores resultados de la validación cruzada en el escenario de 20 factores se obtienen con SVM con acuracy de 94.80 %

y especificidad de 100 %, seguido del modelo híbrido Voting 7 con acuracy de 92.88 %, precisión de 94.82 % y especificidad de 91.08 %.

Tabla 21. *Resultados de la validación cruzada para el escenario de 17 factores con 10-fold*

Modelo	Acu	Pre	Esp
SVM	92.17	86.98	100.0
Perceptron Multilayer	89.90	93.03	87.29
Naive Bayes	73.90	72.91	79.87
Decision Tree	68.04	66.85	73.25
KNN	72.29	72.54	75.96
Random Forest	86.72	91.34	82.50
Gradient Boosting	91.52	93.84	88.88
Voting3	90.58	93.22	88.42
Voting5	92.78	93.47	89.92
Voting7	93.78	93.47	94.92

Fuente. Elaboración propia

En el escenario de 17 factores dado en la Tabla 21, se observa que el mejor resultado de la validación cruzada se obtiene con el modelo híbrido Voting 7 con acuracy de 93.78 %, precisión de 93.47 y especificidad de 94.92 %, seguido de SVM con acuracy de 92.17 %, precisión de 86.98 % y especificidad de 100.00 %.

Tabla 22. *Resultados de la validación cruzada para el escenario de 5 factores con 10-fold*

Modelo	Acu	Pre	Esp
SVM	84.04	84.75	83.21

Perceptron Multilayer	77.23	74.04	84.88
Naive Bayes	68.03	66.00	77.21
Decision Tree	71.00	75.23	63.23
KNN	70.61	74.16	66.42
Random Forest	83.04	86.11	79.08
Gradient Boosting	84.03	83.93	84.04
Voting3	81.44	82.31	79.33
Voting5	84.82	84.34	77.58
Voting7	88.82	84.34	77.58

Fuente: Elaboración propia

En el escenario de 5 factores, en los resultados mostrados en la Tabla 22, se observa que el mejor resultado de la validación cruzada se obtiene con el modelo híbrido Voting 7 con acuracy de 88.82 % y precisión de 84.34 %; seguido de SVM con acuracy de 84.04 %, precisión de 84.75 % y Gradiente Boosting con acuracy de 84.04 % y precisión de 83.93 %. En comparación con los escenarios anteriores, existe una degradación en los resultados, sin embargo, es alcanzado con solo 5 factores dados en la Tabla 19.

Tabla 23. *Matriz de confusión de los resultados de testing para los tres escenarios*

Modelo		Todos los factores (20)		PCA (17)		GreedyStepWise(5)	
SVM	Failure	TN=16	FP=3	TN=17	FP=2	TN=16	FP=3
	Success	FN=0	TP=15	FN=0	TP=15	FN=1	TP=14
Perceptron Multilayer	Failure	TN=19	FP=0	TN=18	FP=1	TN=15	FP=4
	Success	FN=0	TP=15	NP=4	TP=11	FN=2	TP=13
Naive Bayes	Failure	TN=14	FP=5	TN=13	FP=6	TN=11	FP=8
	Success	FN=4	TP=11	FN=5	TP=10	FN=3	TP=12
Decision Tree	Failure	TN=14	FP= 5	TN=10	FP=9	TN=14	FP=5
	Success	FN=6	TP=9	FN=2	TP=13	FN=7	TP=18

KNN	Failure	TN=17	FP=2	TN=18	FP=11	TN=11	FP=8
	Success	FN=2	TP=13	FN=34	TP=11	FN=6	TP=9
Random Forest	Failure	TN=18	FP=1	TN=19	FP=0	TN=16	FP=3
	Success	FN=2	TP=13	FN=3	TP=12	FN=2	TP=13
Gradient Boosting	Failure	TN=19	FP= 0	TN=18	FP= 1	TN=16	FP= 3
	Success	FN=1	TP=14	FN=2	TP=13	FN=2	TP=13
Voting3	Failure	TN=19	FP= 0	TN=18	FP= 1	TN=16	FP= 3
	Success	FN=0	TP=15	FN=2	TP=13	FN=1	TP=14
Voting5	Failure	TN=19	FP= 0	TN=18	FP= 1	TN=16	FP= 3
	Success	FN=0	TP=15	FN=2	TP=13	FN=1	TP=14
Voting7	Failure	TN=19	FP= 0	TN=18	FP= 1	TN=16	FP= 3
	Success	FN=0	TP=15	FN=2	TP=13	FN=1	TP=14

Fuente: Elaboración propia

Por otro lado, la función de pérdida para el entrenamiento y validación de Perceptron Multilayer muestra estabilización en no mayor a 1750 épocas para los cuatro escenarios (ver Figura 33).

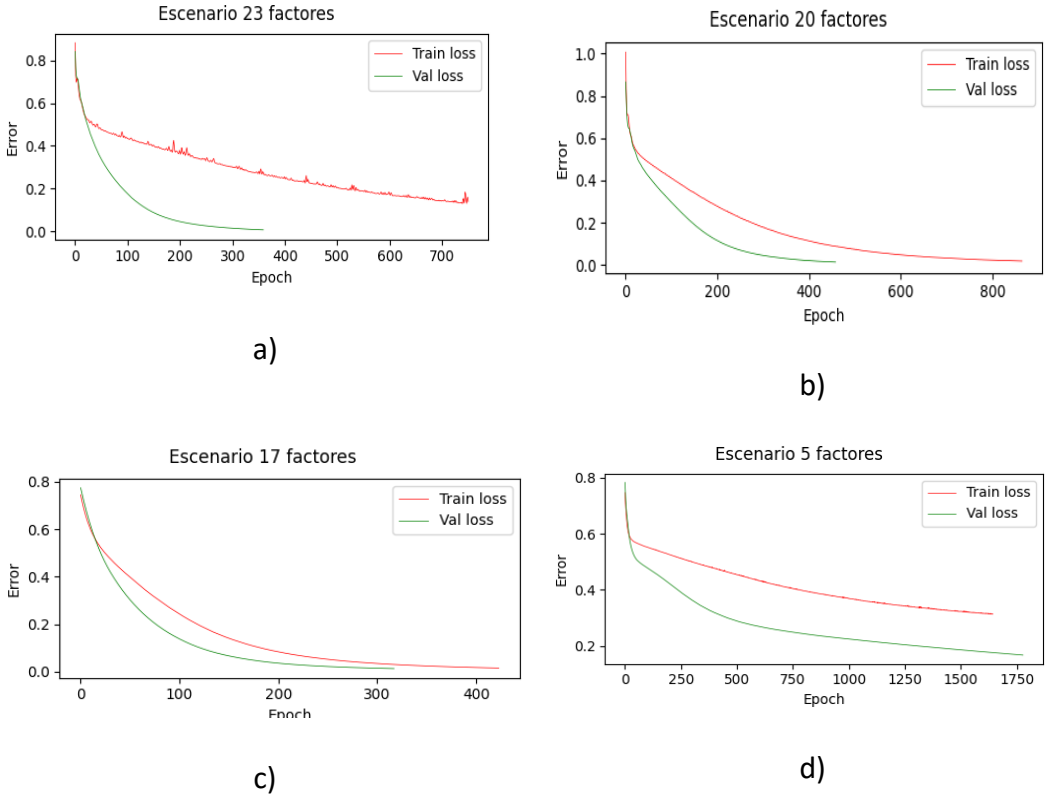

Figura 33. Variación de la pérdida por época del modelo Perceptron Multicapa para el pronóstico del éxito de un STI: a) primer escenario; b) segundo escenario; c) tercer escenario; d) cuarto escenario.

Fuente: Elaboración propia

Los resultados de la *Prueba*, luego del *entrenamiento-validación* para los 7 modelos individuales y los 3 modelos híbridos en los escenarios de 23 factores y 20 factores (sin selección de variables), se muestran en la Tabla 24, así mismo, los escenarios de 17 factores y 5 factores (con selección de variables con PCA y GredyStep Wise) se muestran en la Tabla 25. Los híbridos son dados por la estrategia de votación aplicado a los 7 modelos de ML, a los 5 mejores modelos de ML y a los 3 mejores modelos de ML y que son denotados, respectivamente, por Voting7, Voting5 y Voting3.

Tabla 24. *Resultados de pronóstico del éxito de un STI para 7 modelos de ML y los 3 híbridos en los escenarios de 23 factores y 20 factores.*

Modelos	Primer Escenario (23 factores)			Segundo Escenario (20 factores)		
	Acu	Pre	Esp	Acu	Pre	Esp
Perceptron Multilayer	71.68	77.90	85.50	100.0	100.0	100.0
Gradient Booster	76.00	83.33	83.33	97.06	100.0	95.99
SVM	84.00	84.00	100.0	91.18	83.33	100.0
Random Forest	72.00	77.77	82.35	91.18	92.86	90.00
KNN	64.00	73.68	77.77	82.86	88.24	77.78
Naïve Bayes	68.00	75.00	83.33	73.57	68.65	77.78
Decision Tree	60.00	63.15	80.00	67.65	64.29	70.00
*Promedio**	70.81	76.40	84.61	86.21	85.34	87.36
Voting3	60.00	60.86	93.33	100.0	100.0	100.0

Voting5	64.00	62.50	100.0	100.0	100.0	100.0
Voting7	64.00	62.50	100.0	100.0	100.0	100.0

Fuente: Elaboración propia

Tabla 25. *Resultados de pronóstico del éxito de un STI para 7 modelos de ML y los 3 híbridos en escenarios de 17 factores y 5 factores*

Modelos	Tercer Escenario (17 factores)			Cuarto Escenario (5 factores)		
	Acu	Pre	Esp	Acu	Pre	Esp
Perceptron Multilayer	85.29	91.67	81.82	82.35	76.47	88.24
Gradient Booster	91.18	92.86	90.00	85.29	81.25	88.89
SVM	94.12	88.24	100.0	88.24	82.35	94.12
Random Forest	91.18	100.0	86.36	85.29	81.25	88.89
KNN	55.88	50.00	66.68	58.82	52.94	64.71
Naïve Bayes	67.65	62.50	72.22	67.65	60.00	78.57
Decision Tree	67.65	59.09	83.33	64.71	61.54	66.67
*Promedio**	78.99	77.77	82.92	76.05	70.83	81.44
Voting3	91.18	92.86	86.67	88.24	82.35	93.33
Voting5	91.18	92.86	90.00	88.24	82.35	94.12
Voting7	91.18	92.86	90.00	88.24	82.35	94.12

Fuente: Elaboración propia

De las tablas 24 y 25, se observa lo siguiente:

- El método propuesto en los escenarios 2, 3 y 4 permite obtener modelos de ML con mejores resultados para los 7 algoritmos de ML. En promedio, en el escenario 3, se incrementa la exactitud en 21.75 %, la precisión en 11.69 % y la especificidad en 3.25 %. Estos resultados muestran que las actividades del preprocesamiento, incluyendo el balanceo y la calibración en el proceso de aprendizaje, impactan positivamente en los resultados.
- Los resultados con los 5 factores obtenidos por la aplicación de la rutina *GreedyStepWise* (escenario 4) muestran, respecto al escenario 2, un descenso promedio para los 7 modelos del 12 % para la exactitud, 17 % para la precisión y 7 % para la especificidad. Ello podría explicarse por la pérdida de información generada por la heurística de la rutina, sin embargo, los factores filtrados presentan mucha influencia en el pronóstico, permitiendo obtener con SVM exactitud del 88 %, precisión de 82 % y especificidad de 94 %, lo cual muestra que es muy útil en situaciones donde existan pocos factores estudiados STIs.
- Los modelos individuales que presentan mejores resultados en todos los escenarios son Perceptron Multilayer, Gradient Boosting y SVM, de esta manera, se obtuvo un accuracy de 100 %, 97 % y 91 %, respectivamente. Por otro lado, los 3 modelos híbridos presentan mejores resultados que los modelos de ML por separado; en el segundo escenario, alcanza el valor ideal (100 %) en las tres métricas, incluso en el escenario de 5 factores, en donde se consigue una exactitud del 88 %, precisión arriba del 82 % y especificidad arriba del 93 %.

4.6. Análisis y discusión

Se ha presentado un modelo denominado Information Technology Startup Prediction Model (ITSPM) para predecir el éxito de una Startup de base tecnológica, basada en machine learning, que consta de los siguientes componentes: factores críticos de éxito, extracción de datos, preprocesamiento, modelo predictivo, predicción y resultados. El modelo, es fácil de implementar, dado que existen diversas librerías disponibles y muchas de ellas maduras para cada componente.

Se implementó el modelo – como sistema web - usando los modelos de machine learning SVM, Perceptron Muli-layer, Decision Tree, Naive Bayes, KNN, Radom Forest, Gradient Boosting, así como los modelos híbridos Voting3 (que incluye los 3 mejores modelos respecto a su accuracy), Voting5 (que contiene los 5 mejores modelos respecto a su accuracy) y Voting7 (abarca todos los modelos individuales), además, se realizó el entrenamiento y validación con un data set de 308 registros de startup para cuatro escenarios (23 factores, 20 factores, 17 factores seleccionados por PCA, y 5 factores seleccionados por Greedy-Step-Wise). Los resultados del testing sobre 34 empresas no usadas en el entrenamiento y la validación muestran que los mejores resultados se obtienen para el escenario de 20 factores, alcanzando un accuracy, precisión y especificidad de 100% para Voting3, Voting5, Voting7 y Perceptron Muli-layer. El modelo propuesto es altamente eficaz y permite predecir el éxito de una Startup con alta precisión a pesar de la alta incertidumbre de este tipo de empresa, sin embargo, la precisión del modelo depende de la cantidad y calidad de los datos, factores de éxito contemplados, las actividades de preprocesamiento y los modelos de machine learning considerados.

Por otro lado, los resultados muestran que la reducción de factores usando PCA y Greedy-Step-Wise, en general, reducen el accuracy de los modelos de machine

learning, sin embargo, se alcanza 100 % de precisión con SVM y 17 factores. Además, los resultados obtenidos por Voting5 y Voting7 generan los mejores resultados y se mantienen iguales en los cuatro escenarios, por lo que es suficiente considerar cualquiera de ellos para la predicción del éxito. El algoritmo Greedy-Step-Wise permite identificar los factores Size startup, Company revenue, R&D, Financial capital y Global economic enviroment que influyen fuertemente en el éxito de una Startup, permitiendo alcanzar solo con estos 5 factores un acuracy de 88.24 % para Voting5, Voting7 y SVM. Este resultado muestra que es viable predecir el éxito de una Startup con pocos factores y con alta de precisión.

CAPÍTULO 5: CONCLUSIONES, LIMITACIONES Y TRABAJOS FUTUROS

5.1. Conclusiones

En este trabajo, se ha propuesto un método sistemático basado en algoritmo de aprendizaje automático para construir un modelo predictivo del éxito de un Startup de TI con alta precisión, que consta de 4 procesos (selección de factores críticos de éxito, extracción de datos, preprocesamiento y aprendizaje). A diferencia de otros estudios, que usualmente se centran en una ciudad o región de un país, el método propuesto es sistemático y aplicable a cualquier ciudad o región, además, el estudio contempla un modelo híbrido que la mayoría de las veces proporciona mejores resultados, así como un inventario de 79 factores críticos de éxito.

Para probar la eficiencia del método, este fue aplicado a una base de datos de 265 STI de Australia con 7 algoritmos de aprendizajes (SVM, Perceptron Muli-layer, Decision Tree, Naive Bayes, KNN, Radom Forest y Gradient Boosting) y luego, con los modelos obtenidos, se fue implementado donde Python usó el modelo de pronóstico considerando estos modelos y 3 modelos híbridos basados en la estrategia de Votación. Además, se ha considerado tres escenarios de pruebas, el primero sin aplicar el método, el segundo usando el método y el tercero usando el método, además del algoritmo GreedyStepWise para reducir los factores.

Los resultados muestran que el método propuesto (escenarios 2 y 3) permite obtener modelos de pronóstico con mejores resultados para los 7 algoritmos de aprendizaje usados. En promedio, en el escenario 2, se incrementa la exactitud en 21.75 %, la precisión en 11.69 % y la especificidad en 3.25 %. Estos resultados muestran que los procesos del método impactan positivamente en los resultados. Además, revelan que los mejores resultados se obtienen con el Perceptron Muli-

layer, Gradient Boosting y SVM, con un accuracy de 100 %, 97 % y 91 %, respectivamente. También evidencia que el modelo híbrido, en general, proporciona mejores resultados que los modelos por separado, alcanzando una exactitud ideal del 100 %.

El método propuesto con el algoritmo de GreedyStepWise permite reducir de 20 factores a 5 factores muy significativos (Size startup, Company revenue, R&D, Financial capital y Global economic enviroment) y obtener, a través de SVM y los modelos híbridos, un pronóstico con exactitud de 88 %, precisión de 82 % y especificidad de 94 %, lo cual muestra que es muy útil en situaciones donde existan pocos factores de estudio sobre los STIs.

Los resultados del modelo generado por el método presentan alta precisión, exactitud y especificidad a pesar de la alta incertidumbre de este tipo de empresas, de este modo, se muestra que el método propuesto es sistemático y aplicable a otras realidades. Sin embargo, los resultados dependen de la calidad y cantidad de los datos, los factores de éxito contemplados, las actividades de preprocesamiento y los modelos de aprendizaje automático considerados, es decir, no se pueden extrapolar a otras realidades, pero siguiendo el método se puede garantizar obtener buenos resultados.

5.2. Limitaciones

El dataset usado en los experimentos numéricos es del ecosistema empresarial de Australia, por lo que no podría generalizarse.

5.3. Trabajos futuros

El éxito de una Startup se construye en cada etapa de su ciclo de vida, y cada etapa es dependiente del éxito de las etapas predecesoras, por lo que solo se puede tener éxito en una etapa, si este se alcanza en las etapas previas. Por ejemplo, solo se puede tener éxito en la etapa temprana (etapa donde el modelo de negocio inicial se ha mejorado, y los productos y/o servicios están en el mercado local), si se ha alcanzado éxito en la etapa semilla (etapa donde se inicia la idea innovadora, se pone en marcha el modelo de negocio inicial y se tiene el producto mínimo viable). Por ello, constituye un desafío estudiar la predicción del éxito de una Startup en cada una de las etapas del ciclo de vida.

REFERENCIAS

1. Abakar, K., & Yu, C. (2014). Performance of SVM based on PUK kernel in comparison to SVM based on RBF kernel in prediction of yarn tenacity. Indian Journal of Fibre & Textile Research, 39, 55-59.
2. Albán, M., & Mauricio, D. (2018). Decision Trees for the Early Identification of University Students at Risk of Desertion. International Journal of Engineering & Technology, 7(4.44), 51-54.
3. Almakenzi, S., Bramantoro, A., & Rashideh, W. (2015). A survivability model for Saudi ICT Startups. International Journal of Computer Science & Information Technology, 7(2), 145-157.
4. Alvarez, S., & Barney, J. (2001). How entrepreneurial firms can benefit from alliances with large partners. Acad. Manage. Exec., 15(1), 139-148.
5. Anatolijs, P., Bistrova, J., & Daria, T. (2019). Startup Success Factors in the Capital Attraction Stage: Founders' Perspective. Journal of East-West Business, 25(1), 26-51. doi:10.1080/10669868.2018.1503211
6. Anh, D., Hoa, Q., & Quoc, T. (2012). Critical success factors for Vietnamese software companies: A framework for investigation. Journal of Sociological Research, 3(2), 160-169.
7. Antretter, T., Blohm, I., & Grichnik, D. (2018). Predicting startup survival from digital traces: Towards a procedure for early stage investors. 39th International Conference on Information System, ICIS 2018. San Francisco, United States.
8. Ardito, L., Messeni Petruzelli, A., & Albino, V. (2015). From technological inventions to new products: A systematic review and research agenda of the main enabling factors. European Management Review, 12(3), 113-147.

9. Arruda, C., Silva, V., & Costa, V. (2013). The Brazilian entrepreneurial ecosystem of StartUps: An analysis of entrepreneurship determinants in Brazil as seen from the OECD pillars. Journal of Entrepreneurship and Innovation Managment, 2(3), 17-57.
10. Asmoro, A., Nugroho, L., & Selo. (2018). Prediction Modeling of Software Satartup Success by PLS-SEM Approach. International Journal of Engineering & Technology, 7(4.40), 141-147.
11. Back, D., Clow, K., & Box.T. (1996). Entrepreneurial success: test of a predictive model. Proceedings of the 1996 27th Annual Meeting of the Decision Sciences Institute. Orlando, FL, USA.
12. Baum, J. A., & Silverman, B. S. (2004). Picking Winners or Building Them? Alliance, Intellectual, and Human Capital as Selection Criteria in Venture Financing and Performance of Biotechnology StartUps. Journal of Business Venturing, 19(3), 411–436.
13. Bertoni, F., Colombo, M., & Grilli, L. (2011). Venture capital financing and the growth of high-tech startups: Disentangling treatment from selection effects. Research Policy, 40, 1028-1043.
14. Breiman, L. Random Forests. Machine Learning 45, 5–32 (2001). https://doi.org/10.1023/A:1010933404324
15. Bocken, N. (2015). Sustainable venture capital – catalyst for sustainable startup success? Journal of Cleaner Production, 108, 647-658.
16. Borrajo, L., Baruque, B., Corchado, E., Bajo, J., & Corchado, J. (2011). Hybrid neural intelligent system to predict business failure in small-to-medium-size enterprises. International Journal of Neural Systems, 21(4), 277-296.

17. Cannone, G., & Ughetto, E. (2014). Born globals: A cross-country survey on high-tech startups.,. International Business Review, 23, 272-283.
18. Cerpa, N., & Barden, M., (2016), Evaluating different families of prediction methods for estimating software project outcomes.
19. Chandrashekar, G., & Sahin, F. (2014). A survey on feature selection methods. Computers & Electrical Engineering, 40(1), 16.28.
20. Corrales, D., Ledezma, A., & Corrales, J. C. (2020). A case-based reasoning system for recommendation of data cleaning. Applied Sof Computing Journal, 90.
21. Dellermann, D. L., Ebel, P., Popp, K., Leimeister, & J.M. (2017). Finding the Unicorn: Predicting Early Stage Startup Success through a Hybrid Intelligence Method. International Conference on Information Systems- ICIS. Seoul, South Korea.
22. Diochon, M., Menzies, T., & Gasse, Y. (2007). Attributions and success in new venture creation among Canadian nascent entrepreneurs. Journal of Small Business & Entrepreneurship, 20(4), 335-350.
23. Douzas, G., Bacao, F., & Last, F. (2018). Improving Imbalanced Learning Through a Heuristic Oversampling Method Based on K-Means and SMOTE. Information Sciences, doi:10.1016/j.ins.2018.06.056.
24. Elhedhli, S., Akdemir, C., & Astebro, T. (2014). Classification models via Tabu search: An application to early stage. Expert Systems with Applications, 41, 8085-8091.
25. Freund, Y., y Schapire, R. E. (1997). A decision-theoretic generalization of online learning and an application to boosting. Journal of Computer and System Sciences, 55(1), 119–139

26. Friedman, J. H. (2001). Greedy Function Approximation: A Gradient Boosting Machine. The Annals of Statistics, 29(5), 1189–1232. http://www.jstor.org/stable/2699986
27. Friar, J., & Meyer, M. (2003). Entrepreneurship and startups in the Boston region: Factors differentiating high-growth ventures from micro-ventures. Small Business Economics, 21, 145-152.
28. Ganotakis, P. (2012). Founders' human capital and the performance of UK new technology based firms. Small Business Economics, 39, 495-515.
29. Gartner, W., & Liao, J. (2012). The effects of perceptions of risk, environmental uncertainty, and growth aspirations on new venture creation success. Small Business Economics,, 39, 703-712.
30. Gonzales, A. (2015). Seleccion de variables: Una revision de metodos existentes. Obtenido de http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_1263.pdf
31. Greve, A., & Salaff, J. W. (2003). Social Networks and Entrepreneurship. Entrepreneurship Theory and Practice, 1, 1–20.
32. Grilli, L., & Murtinu, S. (2014). Government, venture capital and the growth of European high-tech entrepreneurial firms. Research Policy,, 43, 1523-1543.
33. Groenewegen, G., & De Langen, F. (2012). Critical success factors of the survival of startups with a radical innovation. Journal of Applied Economics and Business Research, 2(3), 155-171.
34. Haltiwanger, J., Jarmin, R., & Miranda, J. (2012). Who creates obs? Small vs.large vs.young? Unpublished working paper. University of Maryland and US Census Bureau.

35. Helabi, C., & Lussier, R. (2014). A model for predicting small firm performance. Journal of small bussiness and Enterprise Development, 21(1), 4-25.
36. Honorine, A. N., & Emmanuelle, D. (2019). Stage financing and syndication in the IPO underpricing of venture-backed firms: Venture capital and IPO underpricing. The International Journal of Entrepreneurship and Innovation, 20(4), 289-300.
37. Hormiga, E., Batista-Canino, R., & Sánchez-Medina, A. (2010). The role of intellectual capital in the success of new ventures. International Entrepreneurial Management Journal, 1-22.
38. Hormiga, E., Batista-Canino, R., & Sánchez-Medina, A. (2011). The impact of relational capital on the success of new business startups. Journal of Small Business Management, 49(4), 617-638.
39. Hyder, S., & Lussier, R. (2016). Why businesses succeed or fail: A study on small businesses in Pakistan. . Journal of Entrepreneurship in Emerging Economics, 8(1), 82-100.
40. Joshi, K., & Satyanarayana, K. (2014). What ecosystem factors impact the growth of high-tech startups India? . Asian Journal of Innovation and Policy, 3(2), 216-244.
41. Kakadiya, S. M. (2015). Analyzing Startup Success Possibility Using Data Mining. International Journal of Innovations & Advancement in Computer Science, 4(11).
42. kamiran, F., & Calders, T. (2012). Data preprocessing techniques for classification without discrimination. Knowl Inf Syst, 33, 1-33.

43. Kampen, K. J. (2019). Reflections on and test of the metrological properties of summated rating, Likert, and other scales based on sums of ordinal variables. Measurement, 137, 428-434.
44. Kitchenham, B. A.; and Charters, S. (2007). Guidelines for performing systematic literature reviews in software engineering version 2.3. Retrieved November 29, 2017, from http://www.elsevier.com/__data/promis_misc/525444systematicreviewsguide.pdf
45. Krishna, A., Agrawal, A., & Choudhary, A. (2016). Predicting the Outcome of Startups: Less Failure, More Success. International Conference on Data Mining Workshops.
46. Kruchten, P. (1995). Architectural blueprints – The "4+1" View Model of Software Architecture. IEEE Software 12(6), 42-50.
47. Laboissiere, L. A., Fernandes, R. A., & Large, G. G. (2015). Maximum and minimum stock price forecasting of Brazilian power distribution companies based on artificial neural networks. Applied Soft Computing, 35, 66-74.
48. Lasch, F., Le Roy, F., & Yami, S. (2007). Critical growth factors of ICT startups. Management Decision, 45(1), 62-75.
49. Li, S., Shang, J., & Slaughter, A. ((2010).). Why do software companies fail? Information Systems Research, 21(3), 631-654.
50. Maine, E., Shapiro, D., & Vining, A. (2010). The role of clustering in the growth of new technology-based firms. Small Business Economic, 34, 127-146.
51. Martens, D., Vanhoute, C., De Wine, S., Bsesens, B., Sels, L., & Mues, C. (2011). Identifying financial successfull startup profiles with data mining. Expert System with Applications, 38, 5794-5800.

52. Mueller, S., Volery, T., & Von, B. (2012). What do entrepreneurs actually do? An observational study of entrepreneurs everyday behavior in the startup and growth stages. Entrepreneurship Theory and Practice, 995-1017.

53. Mueller, S., Volery, T., & Von, B. (2012). What do entrepreneurs actually do? An observational study of entrepreneurs' everyday behavior in the startup and growth stages. Entrepreneurship Theory and Practic, 995-1017.

54. Nadežda, P., Miroslav, P., & Josef, P. (2019). Factors impacting startup sustainability in the Czech Republic. Innovative Marketing, 15(3), 1-15. doi:10.21511/im.15(3).2019.01

55. OCDE & EUROESTAT, (2006), Manual de Oslo, Guía para la recogida e Reyes OECD (2016). Startup Latina America: building an innovative future Publishing. http://dx.doi.org/10.1787/9789264265141-es. Paris.

56. Pourhashemi, S. M., & Mashalizadeh, A. M. (2013). A novel feature selection method using CFS with Greedy-Stepwise search algorithm in e-mail spam filtering. Advanced Modeling and optimization, 15(3).

57. Pugliese, R., Bortoluzzi, G., & Zupic, I. (2016). Putting process on track: empirical research on startups' growth drivers. Management Decision, 54(7), 1633-1648.

58. Rojas, F., & Huergo, E. (2016). Characteristics of entrepreneurs and public support for NTBFs. Small Business Economics, 1-20.

59. Santisteban, J., & Mauricio, D. (2017). Systematic Literature Review of critical Success Factors of Information Technology Startups. Academy of Entrepreneurship Journal, 23(2).

60. Santisteban, J., Mauricio, D. S., & Cachay, O. (2020). Critical factors success for technology-based satrtups. Article submites for publication.

61. Schneider, J., Dowling, M., & Raghuram, S. (2007). Empowerment as a success factor in startup companies. RMS, 1, 167-184.
62. Sefiani, Y., & Bown, R. (2013). What Influences the Success of Manufacturing SMEs? A Perspective from Tangier. International Journal of Business and Social Science, 4(7), 297-309.
63. Shlens, J. (2014). Obtenido de https://arxiv.org/pdf/1404.1100v1.pdf
64. Singh, D. (2019). Investigating the impact of data normalization on classification performance. Applied Soft Computing, In Press.
65. Song, M., Podoynitsyna, K., Van der Bij, H., & Halman, J. (2008). Success factors in new ventures: A meta-analysis. The Journal of Product Innovation Management, 25, 7-27.
66. Thiranagama, R., & Edirisinghe, K. (2015). Factors affecting small business startup of Engineers and Accountants in Sri Lanka. NSBM Business & Management Journal, 6(1), 84-107.
67. Timmons, J., & Spinelli, S. (2004). New Venture Creation: Entrepreneurship for the 21st Century. New York: McGraw-Hill/Irwin.
68. Tomy, S., & Pardede, E. (2018). From Uncertainties to Successful Start Ups: A Data Analytic Approach to Predict in Technological Entrepreneurship. Sustainability, 10(3), 602.
69. Van de Ven, H., Hudson, R., & Schroeder, M. (1984). Designing new business StartUps. Journal of Management, 10(1), 87-104.
70. Wei-Wen, W. (2009). A competency-based model for the success of an entrepreneurial startup. WSEAS Transactions on Business and Economics, 6(6), 279-291.

71. Yoo, C., Yang, D., Kim, H., & Heo, E. (2012). Key value drivers of StartUp companies in the new media industry – The case of online games in Korea. Journal of Media Economics. 25(4), 244-260.
72. Yoon-Jun, L. (2010). Technology strategy by growth stage of technology-based venture companies. International Review of Business Research Papers, 6(6), 216-234.
73. Yankov, B. & Vitanov, N., (2016), Information System for Forecasting the Success of Bulgarian Start-up Companies.

ANEXOS

ANEXO A

Autovectores generados en la selección de factores con PCA

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	V11	V12	V13	V14	V15	V16	V17	FACTOR
-0.2787	0.0205	-0.1525	0.2034	0.2299	-0.4796	-0.1995	0.0495	-0.16	-0.1151	0.0048	-0.2203	0.1876	-0.3552	0.1315	0.1497	-0.0613	Location
-0.1256	0.2599	-0.2087	-0.0194	0.1756	0.4496	0.1735	-0.3968	0.2316	-0.0945	-0.2439	-0.2227	-0.1606	-0.0096	-0.1009	0.3921	-0.076	Age startup
-0.3596	0.3066	-0.0673	0.201	0.1068	0.0032	-0.1627	-0.038	-0.3135	-0.1457	0.0545	0.0429	-0.0547	0.1451	0.1662	-0.0551	-0.0715	Size_startup
0.0447	0.1553	0.3684	0.3522	-0.3682	0.0532	0.0505	0.2145	0.1957	-0.0112	0.2346	-0.0236	0.1152	-0.0761	0.1973	0.5953	-0.0691	Count profiles skills
-0.3355	0.3914	-0.0593	0.1783	0.0587	0.1335	-0.0567	-0.011	-0.1824	0.0926	0.012	0.1455	-0.0926	0.2214	0.0616	-0.0427	0.1063	Company total revenue
-0.0291	0.1618	0.0814	-0.2108	-0.5572	0.235	-0.3819	-0.0084	-0.3159	-0.0376	-0.1265	-0.1691	0.2817	0.0977	-0.1686	-0.1067	-0.2336	Export products cti
-0.0446	-0.0802	0.0666	0.1454	-0.2964	-0.2218	0.5156	-0.533	-0.2908	-0.2966	-0.02	0.2416	0.1308	-0.0983	-0.0846	-0.0204	-0.0441	Main innvation activities
0.2999	-0.1523	0.0889	-0.0555	-0.0109	0.0635	-0.2909	-0.3143	-0.2694	-0.288	0.1905	-0.4283	-0.3466	-0.1053	0.1627	0.0781	0.2211	Invesment R&D
-0.2012	-0.3892	0.1665	-0.0385	-0.0169	-0.0788	-0.0812	0.1546	-0.2745	0.0166	-0.259	0.1587	-0.4854	0.2076	0.0234	0.3563	-0.082	Domestic economic enviroment
-0.0389	-0.3023	-0.3453	-0.342	0.1225	0.1144	0.109	0.0022	-0.156	-0.0282	0.2978	0.0491	0.3215	0.2908	0.2763	0.2947	-0.3388	availability_of_skilled_employees
-0.3192	-0.117	-0.2088	-0.0369	-0.14	0.0502	-0.0926	-0.0406	0.0484	0.1677	0.5545	0.0109	-0.2144	-0.3294	-0.5243	0.0594	-0.1029	access_to_finance
-0.2825	-0.1023	-0.2382	-0.137	-0.325	0.0038	0.3077	0.193	-0.1325	0.1987	-0.1149	-0.2948	0.0563	0.02	0.07	0.083	0.6288	Cost of_R&D
0.0194	-0.0718	0.4285	0.039	0.3733	0.059	-0.0482	-0.2721	-0.2784	0.476	0.0754	-0.0634	0.3517	0.0979	-0.2505	0.1804	0.2036	Availability infraestructure
-0.2192	0.0536	0.3606	-0.2539	0.0593	0.1362	0.3376	0.0686	-0.122	0.241	-0.0402	-0.2562	-0.1568	-0.3888	0.311	-0.2315	-0.3727	Innovative enviroment
-0.2729	0.0581	0.1666	-0.3526	0.0356	-0.3294	-0.1055	-0.0997	0.3496	-0.2365	-0.1431	-0.2882	0.2248	0.0974	-0.0912	0.1047	0.0124	Goverment regulation
-0.3002	0.0016	0.3884	-0.223	-0.0113	-0.1056	0.0413	-0.0616	0.2028	-0.2003	0.2715	0.0972	-0.1868	0.35	-0.0706	-0.1189	0.0905	Access to target market
-0.1964	-0.3746	0.0901	0.2157	0.0803	0.2374	-0.0508	0.1784	-0.0033	-0.2415	-0.3997	0.0063	0.1383	-0.1417	-0.3308	0.0013	-0.1116	Global economic enviroment
-0.121	-0.2907	0.0331	0.4617	0.0507	0.219	0.148	0.0602	0.0953	-0.1456	0.237	-0.4272	0.1026	0.302	0.0528	-0.3146	-0.0495	Exchange rates
-0.1094	-0.2206	-0.1181	0.2229	-0.2708	-0.2102	-0.216	-0.4263	0.279	0.4863	-0.184	-0.0735	-0.0953	0.1562	0.2045	-0.077	-0.2038	Competitive enviroment
-0.2608	-0.2303	0.0878	-0.0731	-0.0145	0.356	-0.2821	-0.1925	0.1781	-0.1007	0.0447	0.3783	0.1927	-0.3203	0.3958	-0.0797	0.3022	Acces to export markets

ANEXO B

Matriz de correlación generada en la selección de factores con PCA

	F1	F2	F3	F4	F5	F6	F7	F8	F9	F10	F11	F12	F13	F14	F15	F16	F17	F18	F19	F20
F1	1	-0.01	0.48	-0.11	0.28	-0.15	0.02	-0.23	0.14	-0.01	0.27	0.17	-0.04	0.01	0.27	0.09	0.14	0.11	0.14	0.09
F2	-0.01	1	0.27	-0.1	0.37	0	-0.02	-0.15	-0.23	-0.01	0.09	0.07	-0.09	0.08	0.09	-0.01	-0.05	-0.03	-0.06	0.07
F3	0.48	0.27	1	0.02	0.81	0.12	0.02	-0.32	0.02	-0.13	0.26	0.15	-0.06	0.16	0.21	0.24	0.04	0.06	-0.01	0.14
F4	-0.11	-0.1	0.02	1	0.07	0.13	0.07	-0.03	-0.1	-0.39	-0.12	-0.12	0.04	0.05	-0.09	0.07	-0.04	0.11	-0.03	-0.07
F5	0.28	0.37	0.81	0.07	1	0.13	-0.04	-0.45	-0.08	-0.22	0.24	0.21	-0.02	0.18	0.14	0.23	-0.06	-0.01	-0.02	0.1
F6	-0.15	0	0.12	0.13	0.13	1	-0.03	0.07	-0.05	-0.09	0.04	0.1	-0.08	0.05	0.09	0.05	-0.09	-0.21	-0.02	0.07
F7	0.02	-0.02	0.02	0.07	-0.04	-0.03	1	0	0.07	-0.02	0.02	0.12	0	0.04	-0.01	0.09	0.05	0.1	0.12	-0.01
F8	-0.23	-0.15	-0.32	-0.03	-0.45	0.07	0	1	-0.02	0.03	-0.23	-0.32	0.11	-0.2	-0.22	-0.21	-0.1	-0.02	-0.05	-0.09
F9	0.14	-0.23	0.02	-0.1	-0.08	-0.05	0.07	-0.02	1	0.17	0.19	0.23	0.11	0.17	0.12	0.27	0.45	0.19	0.19	0.28
F10	-0.01	-0.01	-0.13	-0.39	-0.22	-0.09	-0.02	0.03	0.17	1	0.2	0.24	-0.09	-0.05	0.01	-0.06	0.09	0.06	0.01	0.17
F11	0.27	0.09	0.26	-0.12	0.24	0.04	0.02	-0.23	0.19	0.2	1	0.37	-0.12	0.1	0.16	0.23	0.18	0.15	0.22	0.31
F12	0.17	0.07	0.15	-0.12	0.21	0.1	0.12	-0.32	0.23	0.24	0.37	1	-0.22	0.2	0.17	0.12	0.15	0.11	0.15	0.14
F13	-0.04	-0.09	-0.06	0.04	-0.02	-0.08	0	0.11	0.11	-0.09	-0.12	-0.22	1	0.21	0	0.12	0.07	0.06	-0.01	0.06
F14	0.01	0.08	0.16	0.05	0.18	0.05	0.04	-0.2	0.17	-0.05	0.1	0.2	0.21	1	0.26	0.38	0.07	-0.02	-0.12	0.16
F15	0.27	0.09	0.21	-0.09	0.14	0.09	-0.01	-0.22	0.12	0.01	0.16	0.17	0	0.26	1	0.52	0.05	-0.13	0.05	0.19
F16	0.09	-0.01	0.24	0.07	0.23	0.05	0.09	-0.21	0.27	-0.06	0.23	0.12	0.12	0.38	0.52	1	0.1	0.06	-0.02	0.29
F17	0.14	-0.05	0.04	-0.04	-0.06	-0.09	0.05	-0.1	0.45	0.09	0.18	0.15	0.07	0.07	0.05	0.1	1	0.43	0.12	0.38
F18	0.11	-0.03	0.06	0.11	-0.01	-0.21	0.1	-0.02	0.19	0.06	0.15	0.11	0.06	-0.02	-0.13	0.06	0.43	1	0.19	0.18
F19	0.14	-0.06	-0.01	-0.03	-0.02	-0.02	0.12	-0.05	0.19	0.01	0.22	0.15	-0.01	-0.12	0.05	-0.02	0.12	0.19	1	0.18
F20	0.09	0.07	0.14	-0.07	0.1	0.07	-0.01	-0.09	0.28	0.17	0.31	0.14	0.06	0.16	0.19	0.29	0.38	0.18	0.18	1

ANEXO C

Factores que influyen en el éxito de una Startup de TI

1. Access to export market	Tomy & Pardede (2018)
2. Access to target market	Tomy & Pardede (2018)
3. Age	Zahra et al (2003), Diochon et al. (2007)
4. Apoyo de incubadora	Santisteban et al., (2020)
5. Apoyo del gobierno	Lasch et al. (2007), Anh et al. (2012), Arruda et al., (2013), Pugliese et al., (2016)
6. Availability of infrastructure	Tomy & Pardede (2018)
7. Availability of skilled employees	Li et al. (2010), Groenewegen & De Langen (2012), Yoo et al (2012), Tomy & Pardede (2018)
8. Business Agent,	Borrajo et al., (2011)
9. Business Model	Böhm et al., (2017)
10. Capacidad dinámica	Santisteban et al., (2020)
11. Capital de riesgo	Bertoni et al. (2011), Grilli & Murtinu, (2014), Bocken (2015), Almakenzi, et al.(2015), Anatolijs et al., (2018)
12. Capital raised	Baum & Silverman, (2004),
13. Clustering	Maine et al (2010), Yoon-Jun, (2010), Mueller et al. (2012)
14. Competition	Song et al. (2008), Arruda et al., (2013), Elhedhli et al., (2014), Tomy & Pardede (2018)
15. Competitive strategy	Asmoro et al., (2018)
16. Cost of Production	Elhedhli et al., (2014)
17. Cultura innovadora	Santisteban et al., (2020)
18. Ecosistema de innovación y emprendimiento	Santisteban et al., (2020)
19. Edad del negocio	Haltiwanger et al. (2012)
20. Education	Backet al., (1996), Hyder & Lussier, (2016), Pugliese et al., (2016), Rojas & Huergo, (2016), Helabi & Lussier, (2014)
21. Entrepreneurial education	Baum & Silverman(2004), Maxwell, et al. (2011)
22. Entrepreneurial experience	Gartner & Liao (2012), Mueller et al.., (2012),Bocken (2015), Pugliese et al., (2016)
23. Environment	Martens et al., (2011), Asmoro et al., (2018), Timmons & Spinelli (2004), Tomy & Pardede (2018)
24. Evaluator Agent	Borrajo et al., (2011)

25. Exchange rates	Tomy & Pardede (2018)
26. Experiencia en gestión de empresas	Back et al., (1996), Arruda et al., (2013), Cannone & Ughetto (2014), Thiranagama & Edirisinghe (2015), Hyder & Lussier (2016)
27. Expert Agent	Borrajo et al., (2011)
28. Financial and accouting Information	Helabi & Lussier, (2014)
29. Financial capital	Martens et al., (2011), Tomy & Pardede (2018)
30. Financiamiento por etapas	Santisteban et al., (2020)
31. Founder	Schneider et al. (2007), Wei-Wen, (2009), Asmoro et al., (2018)
32. Functional Performance	Elhedhli et al., (2014)
33. Género del emprendedor	Friar & Meyer, (2003)
34. Government regulation	Tomy & Pardede (2018), Pugliese et al. (2016)
35. Human capital	Martens et al., (2011)
36. Industry	Thiranagama & Edirisinghe, (2015); Hyder & Lussier, (2016); Pugliese et al., (2016); Rojas & Huergo (2016)
37. Innovación de producto /servicio	Ardito et al., (2015)
38. Innovation environment	Tomy & Pardede (2018)
39. Internet	Helabi & Lussier, (2014)
40. Knowledge support	Maxwell, et al. (2011)
41. Marketing	Helabi & Lussier, (2014)
42. Motivation	Greve & Salaff (2003), Ganotakis (2012)
43. Need	Elhedhli et al., (2014)
44. Organization	Martens et al., (2011), Asmoro et al., (2018)
45. Partners	Sefiani & Bown(2013), Helabi & Lussier, (2014)
46. Planning	Helabi & Lussier, (2014)
47. Potential Market	Elhedhli et al., (2014)
48. Price	Elhedhli et al., (2014)
49. Profitability	Elhedhli et al., (2014)
50. Proof of Concept	Maxwel, et al., (2011)
51. Proof of value	Shepherd & Zacharakis(1999); Mason & Stark (2004)

52. Research and Development	Baum & Silverman, (2004), Elhedhli et al., (2014), Tomy & Pardede (2018)
53. Resource	Asmoro et al., (2018)
54. Company revenue	Böhm et al. (2017)
55.Satisfacción del cliente	Santisteban et al., (2020)
56. Service	Elhedhli et al., (2014)
57. Size of Investment	Elhedhli et al., (2014)
58. Social capital	Martens et al., (2011)
59. Store Agent	Borrajo et al., (2011)
60. Size startup	Joshi & Satyanarayana, (2014), Cannone & Ughetto (2014), Thiranagama & Edirisinghe (2015), Rojas & Huergo (2016)
61. Technical Feasibility	Elhedhli et al., (2014)
62. Technological Hype	Maxwell et. al (2011)
63. Technology Significance.	Elhedhli et al., (2014)
64. Location	Hormiga et al. (2011)
65. Value Creation Process	Asmoro et al., (2018)
66. Web analitics	Silver (2012)
67. Working capital	Helabi & Lussier, (2014)
68. Capacidades tecnológicas / empresariales	Li et al., 2010; Groenewegen & De Langen ,2012; Yoo et al., 2012
69. Global economic environment	Tomy & Pardede, 2018
70. Technological surveillance	Ko & An (2019), Roa et al. (2018)
71. Knowledge absorptive capacity	(Senivongse et al., 2019)
72. Perceived performance	(Arefin et al., 2019)
73. Quality of a product and/or service	(Al-Fraihat et al., 2020)
74. Customer satisfaction	Luna-Perejon et al., 2019
75. Staged financing	Honorine & Emmanuelle (2019)
76 Support of a business incubator	Murray (2019)
77. Innovation and entrepreneurship ecosystem	Corrales-Estrada (2019)
78. Dynamic capability of entrepreneurs	Arora et al., 2019
79. Innovative and entrepreneurial culture	Roy et al. (2020), Corrales-Estrada (2019)

Se ha considerado los sinónimos como sigue. *Entrepreneurial experience* (Maxwel et al., 2011), experiencia previa en puesta en marcha (Gartner & Liao, 2012; Mueller et al., 2012: Bocken, 2015; Pugliese et al., 2016); *Education* (Helabi & Lussier, 2014), formación académica (Hyder & Lussier, 2016; Pugliese et al. 2016; Rojas & Huergo, 2016); *Availability of skilled employees* (Tomy & Pardede, 2018), Access skill; *capacidades tecnológicas/empresariales del equipo* (Li et al., 2010; Groenewegen & De Langen, 2012; Yoo et al., 2012), Profile skills; *R&D* (Elhedhli et al., 2014), Research and Development (Tomy & Pardede, 2018), experiencia en investigación y desarrollo (Baum & Silverman, 2004); *Potential market* (Elhedhli et al., (2014), exports CTI products; *Innovación de producto/servicio* (Ardito et al., (2015), Main innovation activities; *Size of investment* (Elhedhli et al., 2014), inversión en R&D; *Founder* (Asmoro et al., 2018), liderazgo del emprendedor fundador (Schneider et al., 2007; Wei-Wen, 2009); *Age* (Zahra et al., 2003), edad del líder emprendedor (Diochon et al., 2007); *Competition* (Landström,1998), competitive environment (Tomy & Pardede, 2018), new competition (Elhedhli et al., 2014), mercado competidor (Song et al., 2008; Arruda et al., 2013); *Size startup*(Joshi & Satyanarayana, 2014; Cannone & Ughetto, 2014; Thiranagama & Edirisinghe, 2015; Rojas & Huergo, 2016), Team size (Edirisinghe, 2015), team constellation (Kirsch et al., 2009); *Environment* (Asmoro et al., 2018), Domestic economic environment (Tomy & Pardede, 2018), dinamismo del entorno (Timmons & Spinelli, 2004); *Government regulation* (Tomy & Pardede, 2018), políticas en ciencia y tecnología (Pugliese et al., 2016); *Organization* (Asmoro et al., 2018), capital organizacional (Martens et al., 2011), y *capital financiero* (Martens et al., 2011), Financial support (Baum & Silverman, 2004), Access to finance (Tomy & Pardede, 2018).

ANEXO D

Producción científica de la investigación

International Journal of Information Technology and Web Engineering
Volume 18 • Issue 1

Predicting the Success of a Startup in Information Technology Through Machine Learning

Edilberto Vasquez, AI Group, Universidad Nacional Mayor de San Marcos, Peru
José Santisteban, AI Group, Universidad Nacional Mayor de San Marcos, Peru*
David Mauricio, AI Group, Universidad Nacional Mayor de San Marcos, Peru
https://orcid.org/0000-0001-9262-626X

ABSTRACT

Predicting the success of a startup in information technology (SIT) is a very complex problem due to the diverse factors and uncertainty that affects it. The focus of automatic learning (ML) is promising because it presents good results for prediction issues; however, it presents a diversity of parameters, factors, and data that require consideration to improve prediction results. In this study, a systematic method is proposed to build a predictive model for SIT success, based on factors. The method consists of four processes, a hybrid model, and an inventory of 79 success factors. The method was applied to a database of 265 SITs from Australia with seven ML algorithms and three hybrid models based on the Voting strategy and the GreedyStepwise algorithm to reduce the factors. On average, precision increments in 11.69%, specificity in 3.25%, and accuracy in 21.75%; the prediction has precision of 82% and accuracy of 88%.

KEYWORDS

Critical Success Factors, Forecast, Machine Learning, Startups

INTRODUCTION

Predicting the Success of a Startup in Information Technology Through Machine Learning

A technology-based startup is defined as the grouping of people around an innovative technology-based idea with a replicable and scalable business model (Nadežda et al., 2019); it is an innovative venture that provides solutions to emerging problems or creates new demands by developing new forms of business (OECD, 2005). It is widely established that entrepreneurship is important for the wealth and economic growth of countries (Cabrera & Mauricio, 2017). In this regard, the importance of startups in information technology (SITs) lies in the revitalization of economies, directly impacting the creation of jobs, products and/or services with high added value. Moreover, various World Bank studies show that emerging technological companies, in 2017, contributed more than 5% of the gross

DOI: 10.4018/IJITWE.323657
*Corresponding Author