

Big Data and Machine Learning for Applied Economics – MECA 4107 – 2022-13

Maestría en Economía Aplicada – Universidad de los Andes

Proyecto final – Equipo 12

Ingrid L. Molano, 200511102

Jorge E. García, 201310645

Camilo Villa, 201818624

Agosto 2, 2022

“Clasificación de propuestas de investigación como financiables o no financiables mediante la predicción de un índice de impacto científico¹ en los Programas Marco de Investigación e Innovación de la Comisión Europea, con base en características del consorcio² y sus integrantes”

Repositorio: https://github.com/camilovillam/MECA_BD_Final_project

Resumen:

Este trabajo tiene como propósito desarrollar un modelo predictivo para determinar si una nueva propuesta de investigación será exitosa o no al momento de aplicar por financiación en los Programas Marco de Investigación e Innovación de la Comisión Europea. Para esto, se toma como base el estudio de Wanzenböck, Lata, Ince (2020), y se replica con la información disponible en CORDIS³. Uno de los grandes retos de este trabajo fue el no tener acceso a la información de los proyectos no financiados. Para mitigar esto, se simuló una base de prueba generando consorcios aleatorios que cumplen con las condiciones básicas de participación en estos programas de la comisión europea. En particular se aplican modelos de predicción basados en algoritmos de aprendizaje de máquinas y estimación bajo validación cruzada para variables que identifican las características del consorcio y el proyecto de investigación, luego se calcula el índice científico del proyecto y se aplica el modelo de clasificación.

¹ Se entiende por índice de impacto científico el efecto producido por el desarrollo del proyecto de investigación (Número de publicaciones, patentes, entre otros).

² Un consorcio es una unión temporal entre instituciones (universidades, centros de investigación, empresas, entidades públicas) para presentar propuestas y llevar a cabo proyectos de investigación.

³ Servicio de Información Comunitario sobre Investigación y Desarrollo de la Comunidad Europea. En particular se utilizó la información de los proyectos financiados en programas marco FP7 y Horizon 2020

Introducción:

La investigación y la innovación dependen cada vez más de equipos y consorcios de investigadores e instituciones (Wang, Barabási, 2021). Los Programas Marco de Investigación e Innovación de la Comisión Europea (*Framework Programme FPI* hasta el *FP7*, *Horizon 2020*, *Horizon Europe*) son uno de los programas de investigación e innovación más grandes e importantes en el mundo. Si bien tienen un énfasis en Europa, las convocatorias están abiertas a la participación de instituciones de otros países, como Colombia, en consorcios.

Las convocatorias de los programas marco europeos son altamente competitivas y las exigencias tanto en términos de la conformación de los consorcios como de la calidad científica de las propuestas son muy altos. Conformar un consorcio potencialmente ganador y preparar una propuesta competitiva es complejo, exige una dedicación importante de investigadores y personal administrativo, y requiere un conocimiento detallado de los términos y condiciones de las convocatorias. Por ello, tener la capacidad de preparar y presentar propuestas exitosas en este tipo de convocatorias internacionales está en el corazón de la productividad científica y de la economía de la ciencia.

Existen diferentes estudios que demuestran que el éxito de una propuesta en una convocatoria depende en gran medida de la conformación del consorcio, de la experiencia y capacidad del coordinador, del nivel de productividad científica⁴ de las instituciones que lo conforman; de si ya han trabajado antes juntos, entre otros factores (Wanzenböck, Lata, Ince, 2020). Teniendo en cuenta este contexto, es relevante para instituciones científicas, como las universidades, tener mejores herramientas para hacer propuestas más competitivas que representen una mayor tasa de éxito y mayores recursos externos para sus actividades de investigación e innovación.

Partiendo de un artículo que presenta un análisis de factores de éxito en propuestas en proyectos de Horizon 2020, (Wanzenböck, Lata, Ince, 2020), este trabajo tiene como propósito **desarrollar un modelo predictivo basado en aprendizaje de máquinas para determinar si una nueva propuesta será exitosa o no, con base en características del consorcio y sus integrantes**. Es decir, el foco está en la predicción de la probabilidad de éxito de propuestas nuevas, fuera de muestra. Es decir, el valor agregado de esta propuesta es hacer una predicción basada en el análisis del artículo, pero usando datos abiertos que, a diferencia de los utilizados por los autores en dicho artículo, están disponibles para cualquier institución.

En el desarrollo de este trabajo, se aplicaron diferentes modelos de predicción para las variables de interés requeridas para el cálculo del índice de impacto científico. Se seleccionó para cada variable el modelo de predicción con menor error cuadrático medio y, se procedió con la aplicación del modelo de clasificación para determinar si un proyecto de investigación es financiable o no. Una vez hecho este ejercicio se encontró que, sobre la base de prueba de la base de entrenamiento el 98% de los proyectos clasificaron como financiables, dejando solo un 2% como falsos negativos.

Lo anterior, se considera como un resultado positivo dado que la información base sobre la que se llevaron a cabo las predicciones era de proyectos que fueron financiados. El gran reto fue determinar si el modelo aplicaría a proyectos que no cumplan las características para ser financiados, clasificándolos como no financiables. Para contrarrestar esto, se simuló una base de prueba de proyectos ficticios.

La aplicación del modelo de clasificación sobre esta base simulada arrojó que el 19% de los proyectos/consorcios simulados no son financiables. Lo anterior podría interpretarse como un resultado favorable en la medida que se esperaba que la simulación aleatoria hiciera que el resultado fuera que no todos los proyectos fueran financiables. No obstante, se considera que esto no es suficiente ya que, los modelos de predicción aplicados y el modelo de clasificación desarrollado pueden presentar aún un sesgo por haber sido entrenados en una base que no permitía observar características de proyectos no financiables. En este sentido, y pensando en un trabajo futuro con la intención de seguir mejorando el modelo, se considera necesario replicar el ejercicio desarrollado en este trabajo

⁴ Se entiende por productividad científica el número de investigaciones publicadas de una universidad.

con una base de entrenamiento que incluya una muestra representativa de proyectos financiados y proyectos no financiados.

Datos:

El artículo de Wanzenböck, Lata e Ince (2020) emplea como fuente de información la base de datos *CORDA - Common Research Datawarehouse* de la Comunidad Europea, y el subconjunto eCORDA – External CORDA. Estas bases de datos no están disponibles para el público general sino para tomadores de decisión y ciertos analistas en Europa, ya que incluyen alguna información clasificada. Por lo tanto, no es posible replicar el mismo estudio usando los mismos datos. Ahora bien, la información que sí está disponible de manera abierta proviene de CORDIS - Servicio de Información Comunitario sobre Investigación y Desarrollo de la Comunidad Europea. CORDIS tiene información abierta de más de 110.000 proyectos **financiados**, con registros de 1994 en adelante que cubren todos los programas marco (desde el *FP1* hasta *Horizon 2020* y el actual *Horizon Europe*)⁵.

El hecho que CORDIS no tenga información de las propuestas rechazadas, como sí es el caso de CORDA, presenta un gran reto para el objetivo de este trabajo, en la medida que no se pueden observar las características de los consorcios y propuestas de investigación rechazados. Para mitigar esto, se generó una base de proyectos aleatoria a partir de la información contenida en *Horizon 2020*, de tal forma que se crearon consorcios y, por ende, proyectos de investigación ficticios, cumpliendo las reglas básicas de conformación de los consorcios que aplican a estos programas de financiación, que son como mínimo tres organizaciones europeas de tres países europeos distintos. El objetivo de ese ejercicio fue generar una base que tuviera en alguna medida proyectos con características no financiados.

Se resalta que dentro de la información disponible en CORDIS se encuentra: listado de proyectos financiados; convocatoria específica (dentro de cada programa marco hay diferentes subprogramas y tipos de convocatorias); monto total del proyecto; monto co-financiado por la Comisión Europea; instituciones integrantes del consorcio; roles de los socios (ej. coordinador, socio); tipo de institución (universidad, centro de investigación, empresa); detalle de publicaciones derivadas del proyecto; detalle de patentes derivadas del proyecto; otros entregables; resumen del proyecto y sus resultados finales; palabras clave por temas, entre otros.

Tabla 1. Información de organizaciones y Proyectos (elaboración propia con información CORDIS).

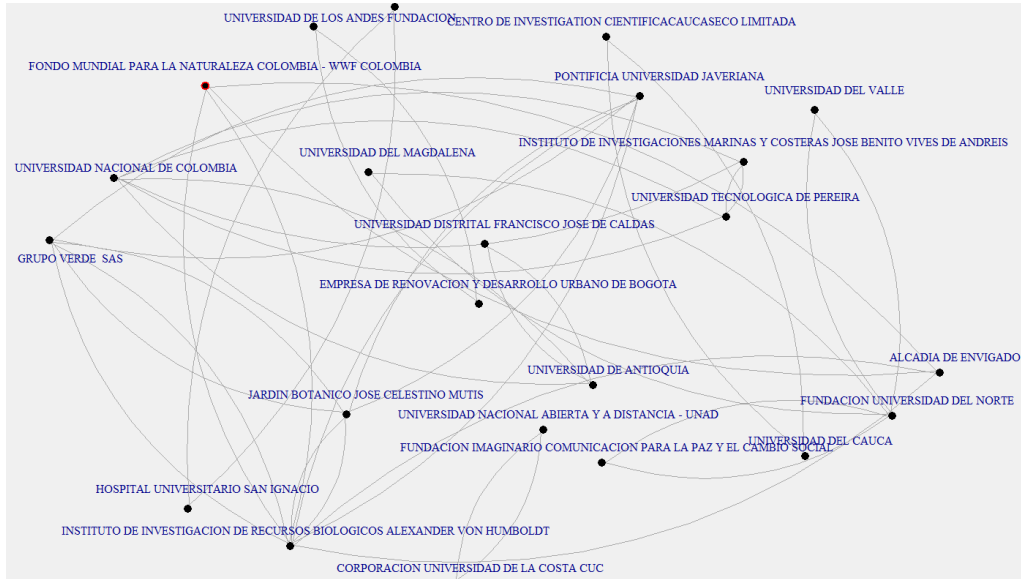
Organizaciones		
Organizaciones totales	40.735	
Organizaciones EU 13	4.470	10,97%
Organizaciones EU 15	29.775	73,09%
Organizaciones no europeas	6.490	15,93%
*organizaciones colombianas	56	0,14%
Proyectos		
Proyectos totales (consorcios) en H2020 - base de entrenamiento	35.352	
Proyectos ficticios - base de prueba	11.415	
Costo Total* (Del costo total se puede financiar entre el 80% y 100%)	Mínimo	\$ 3.938.000
	Mediana	\$ 998.400.000
	Promedio	\$ 2.366.000.000
	Máximo	\$ 1.330.000.000.000

A partir de esta información, es posible determinar variables adicionales como la experiencia de la institución en programas previos; la experiencia de la organización como coordinador o socio, y la experiencia de trabajo previo entre los mismos miembros del consorcio. En el anexo 5 se pueden ver las estadísticas descriptivas de las variables mencionadas.

⁵ CORDIS - EU research projects under Horizon 2020 (2014-2020): <https://data.europa.eu/data/datasets/cordish2020projects?locale=en>
CORDIS - EU research projects under FP7 (2007-2013): <https://data.europa.eu/data/datasets/cordisfp7projects?locale=en>
CORDIS reference data: <https://data.europa.eu/data/datasets/cordisref-data?locale=en>

Por otro lado, se destaca que dentro de las variables utilizadas en el estudio de Wanzenböck, Lata, Ince (2020) y, que a su vez se utilizan en este trabajo, son aquellas que identifican las propiedades de las organizaciones en el marco de la red de los consorcios formados para los proyectos de investigación. En particular, las variables utilizadas son el grado y el vector propio de centralidad dentro de cada organización dentro de la red. La figura 1 ilustra, a modo de ejemplo, una de estas redes, en este caso, las de instituciones colombianas participantes en este tipo de proyectos.

Figura 1. Red de organizaciones colombianas en H2020 (elaboración propia).



Modelo:

Metodología para la construcción del modelo de clasificación:

Para la construcción del modelo de predicción de clasificación de propuestas se siguieron los siguientes pasos:

1. **Estimación / predicción:** con base en características del consorcio y los socios, así como de la convocatoria y programa específico (variables X), se entrenaron modelos y algoritmos para predecir las siguientes variables de resultado: número de publicaciones; número de patentes; otros productos de conocimiento; tamaño del proyecto en euros. Esta información está disponible para la base de entrenamiento, y se predice por fuera de la muestra
2. **Cálculo de un índice de impacto científico:** con estas variables de resultado, se calculó un índice ponderado de impacto científico de los proyectos. El índice se calculó con la información disponible en la base de entrenamiento, mientras que para las observaciones por fuera de muestra se predijo.
3. **Clasificación:** dado que todos los proyectos de la base de datos fueron financiados, el índice calculado para los proyectos de dicha base fue el rango del índice para los proyectos financiados. Cualquier valor del índice predicho por debajo del umbral fue clasificado como “No financiable”; cualquier valor igual o por encima sería “financiable”.

Selección y estimación del modelo de predicción:

Para cada una de las variables a predecir se aplicaron cinco algoritmos diferentes de predicción que son OLS con validación cruzada K-fold = 5, lasso, ridge, elastic net y bosques aleatorios de predicción. Para cada variable de interés se seleccionó el modelo con menor error cuadrático medio (MSE). La tabla completa comparativa de todos los modelos estimados se presenta en los anexos. Todos los modelos emplearon las siguientes variables independientes: tamaño del consorcio; tipo de convocatoria; proporción de universidades en el consorcio; proporción de centros de investigación; proporción de instituciones de países de los grupo de EU-15, EU-13 y no

Europeos; familiaridad y trabajo previo de los miembros del consorcio; el eigenvector de la red de instituciones; la experiencia del coordinador en proyectos FP7; el país del coordinador; la posición del coordinador en el ranking CWTS Leiden 2022; evidencia de patentes otorgadas al coordinador; la cantidad de instituciones del consorcio presentes en el top 50 del ranking CWTS Leiden 2022. En el anexo se incluye una tabla con una descripción de estas variables.

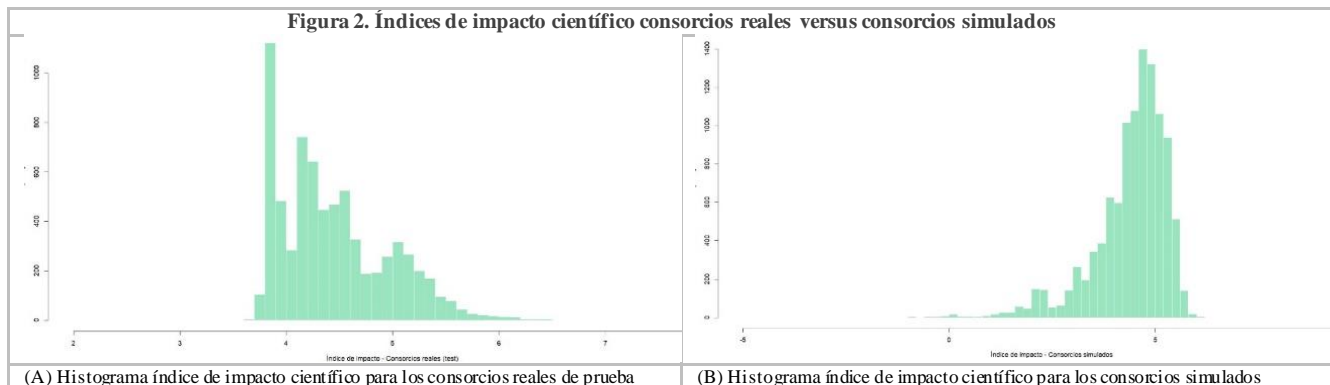
Tabla 2. Mejor modelo de predicción por variable de interés (elaboración propia).

Variable	Mejor modelo	MSE mejor modelo
Contribución de la Comisión Europea (recursos en euros)	<i>Random Forest</i>	1,06315E+13
Logaritmo natural de la contribución de la Comisión Europea	<i>Random Forest</i>	0,534392431
Logaritmo natural del costo total del proyecto	<i>Random Forest</i>	0,536588374
Número de entregables total del proyecto*	<i>Random Forest</i>	48,48583484
Número de artículos <i>peer-reviewed</i>	<i>Ridge</i>	251,449512
Otras publicaciones**	<i>OLS</i>	110,5575862
Número de publicaciones totales	<i>Ridge</i>	455,5797779
Número de patentes	<i>Lasso</i>	1,633076751
Costo total del proyecto	<i>Random Forest</i>	3,64132E+13

* Reportes, prototipos, demostraciones, páginas web, set de datos de *Open Research*

**Todas las demás publicaciones excluyendo los artículos de pares: libros, capítulos de libros, disertaciones de tesis, artículos sin revisión de pares, etc.

Con base en los mejores modelos se realizó la predicción para cada variable de impacto. Y, tal como se describió previamente, a partir de estas predicciones se calculó un índice de impacto del proyecto, que es una variable de impacto potencial de los proyectos, basada en los entregables y resultados predichos. En el anexo se presenta la fórmula de cálculo de dicho índice. Teniendo en cuenta que todos los proyectos de nuestra base de datos fueron financiados, se tomó como umbral inferior un índice de 3.8. Es decir, cualquier valor predicho por debajo de 3.8 representaría clasificar un consorcio como no financiable. Esto se ilustra en los histogramas de la figura 2 para la base de test simulada y la extraída de los proyectos reales.



Resultados:

Al aplicar el modelo de clasificación a la base de prueba de la base de entrenamiento se obtienen un 98% de verdaderos positivos, proyectos que son financiados y son clasificados como financiables, y tan solo el 2% se clasifica como proyectos no financiables (falsos negativos). Al aplicar el modelo de clasificación a la base de prueba simulada con proyectos ficticios aleatorios se obtiene aproximadamente el 81% de los proyectos son financiados y el restante 19% no son financiados. En este punto se desconoce la realidad. Sin embargo, la simulación aleatoria de una base de prueba se realizó con el objetivo de predecir resultados sobre una base de proyectos que no fuera completamente financiable, por eso el componente de aleatoriedad al momento de simularla.

Tabla 3. Matriz de confusión base de prueba de la base de entrenamiento (elaboración propia).

Valores predichos	Valores reales		
	Proyectos Financiados	Proyectos no financiados	
	Proyectos Financiados		
	6.958	0	6.958
	110	0	110
	7.068	0	

En principio, estos resultados son de carácter preliminar y no permiten ni descartar la propuesta metodológica, ni confirmarla. Puede ser relevante conseguir una muestra de proyectos reales no financiados, así sea pequeña, para revisar la capacidad predictiva real. Si bien ésta es información clasificada, sería eventualmente posible conseguirla con instituciones aliadas o a través de redes internacionales de colaboración.

Conclusiones y lecciones aprendidas:

El presente trabajo buscaba aportar a la solución de un problema práctico de las universidades relacionado con el esfuerzo requerido para presentar propuestas de investigación a convocatorias internacionales con muchos recursos pero altamente competitivas, asociado también a una tasa de éxito baja. Con algoritmos y modelos de *machine learning*, se buscaba desarrollar un modelo de recomendaciones para formar consorcios con buen potencial de éxito. Los resultados obtenidos son preliminares pero permiten en principio continuar desarrollando este trabajo y refinar los modelos.

Es importante además dejar explícitas diferentes limitaciones del trabajo aquí desarrollado. La principal de ellas, ya mencionada, tiene que ver con el hecho de la no disponibilidad de observaciones reales de proyectos no financiados. Es decir, estamos ante un caso de *survivor bias*, en el que solo podemos conocer la información de los proyectos exitosos. En esta línea, no tenemos certeza de si los proyectos rechazados fueron descartados justamente por variables inobservables para nosotros (pero sí para los evaluadores) u otros factores no presentes en los exitosos. Esta limitación no es menor, ya que inevitablemente introduce sesgos en las estimaciones y entrenamiento de los modelos, y no nos permite determinar con certeza la capacidad predictiva real de nuestros modelos.

Otra limitación importante tiene que ver con el hecho de la dificultad para emparejar bases de datos de orígenes muy distintos, como es el caso de bases de datos de patentes internacionales con las de los proyectos de investigación europeos. El emparejamiento por nombre de institución es poco exitoso y es un problema conocido en la ciencia (*name & affiliation disambiguation*), debido a diferentes formas de nombrar a las instituciones. Corresponde entonces intentar otros métodos de emparejamiento como la georreferenciación y el tratamiento inteligente de texto, para lo cual los métodos de aprendizaje de máquinas aprendidos resultan muy útiles. Un primer intento de georreferenciación mostró resultados prometedores, aunque éstos deben ser tratados con cautela y validados con otros métodos de emparejamiento, ya que puede inducir a imprecisiones. Esta limitación es más de índole práctico y puede ser superada con un tratamiento adicional de la información, aunque no trivial. También es posible continuar ajustando los modelos, intentando realizar predicciones diferenciadas por tipos de proyectos, los tipos de convocatorias específicas y las temáticas de las mismas, que son líneas que por el alcance no logramos incluir en este análisis pero que pueden ser prometedoras.

En cualquier caso, como con cualquier ejercicio de *big data* y *machine learning*, los resultados del modelo, de lograrse en efecto, deben ser tratados con cautela y a lo sumo deberían tomarse como recomendaciones para fortalecer las propuestas y consorcios de investigación, pero nunca para desestimular de entrada la elaboración y presentación de propuestas a convocatorias internacionales.

Referencias:

Glossary of Statistical Terms (sf). Statistics Portal. En OECD. Recuperado en 02 de agosto de 2022, de: <https://stats.oecd.org/glossary/detail.asp?ID=6805>

Harrap, N., & Doussineau, M. (2017). *Collaboration And Networks: Eu13 Participation in International Science. Stairway To Excellence - Jrc Policy Insights*. Recuperado de: <https://Joint-Research-Centre.Ec.Europa.Eu/System/Files/2017-03/Jrc106208.Pdf>

Wang, D., & Barabási, A. (2021). *The Science of Science*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/9781108610834>.

Wanzenböck, I., Lata, R., & Ince, D. (2020). *Proposal success in Horizon 2020: A study of the influence of consortium characteristics*. Quantitative Science Studies, 1(3), 1136–1158. https://doi.org/10.1162/qss_a_00067

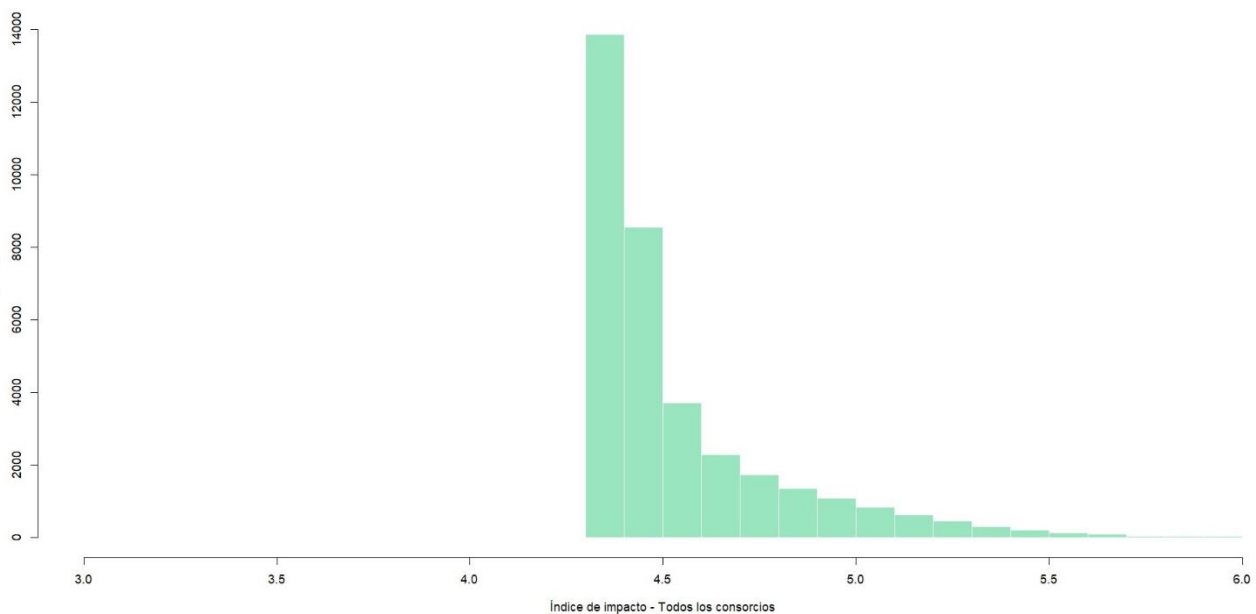
Anexos:

Anexo 1. Modelos de predicción por variable de interés

Modelo por variable de interés	MSE
Contribución de la Comisión Europea	
<i>Random Forest</i>	1,0631E+13
<i>Elastic net</i>	1,1044E+13
<i>Ridge</i>	1,1534E+13
<i>OLS</i>	1,2121E+13
<i>Lasso</i>	1,2123E+13
Logaritmo natural de la contribución de la Comisión Europea	
<i>Random Forest</i>	0,53439243
<i>OLS</i>	0,60943048
<i>Elastic net</i>	0,60948277
<i>Lasso</i>	0,61216215
<i>Ridge</i>	0,61536659
Logaritmo natural del costo total del proyecto	
<i>Random Forest</i>	0,53658837
<i>OLS</i>	0,60697088
<i>Elastic net</i>	0,60881716
<i>Lasso</i>	0,61281801
<i>Ridge</i>	0,61862032
Número de entregables total del proyecto	
<i>Random Forest</i>	48,4858348
<i>Lasso</i>	50,4139613
<i>Ridge</i>	50,418533
<i>Elastic net</i>	50,4562229
<i>OLS</i>	50,4982895
Número de artículos peer-reviewed	
<i>Ridge</i>	251,449512
<i>Lasso</i>	262,32944
<i>Elastic net</i>	412,694374
<i>Random Forest</i>	521,596111
<i>OLS</i>	535,870549

Otras publicaciones	
OLS	110,557586
Random Forest	110,735387
Elastic net	111,24974
Ridge	111,36553
Lasso	111,400464
Número de publicaciones totales	
Ridge	455,579778
Lasso	483,569612
Elastic net	666,643524
OLS	748,343884
Random Forest	761,249549
Número de patentes	
Lasso	1,63307675
Ridge	1,6332538
Elastic net	1,65923177
OLS	1,66972612
Random Forest	1,67044047
Costo total del proyecto	
Random Forest	3,6413E+13
Elastic net	4,0247E+13
Ridge	4,2148E+13
Lasso	4,2971E+13
OLS	4,3026E+13

Anexo 2. Histograma índice de impacto científico para todos los consorcios



Anexo 3. Estadísticas Descriptivas

Nombre de variable	Descripción	Mínimo	Máximo	Media	Mediana	Desviación estándar
ecMaxContribution	contribución (\$) de la Comisión Europea	3150	678800000	1932656	914859	5535986
consorc_size	Tamaño del consorcio	1.000	210	4.93	1.000	8.24
numPartnPRC	Número de socios asignados al sector privado	0	101	1.643	0	3.76

Nombre de variable	Descripción	Mínimo	Máximo	Media	Mediana	Desviación estándar
total_Cost	Costo total del proyecto medido en euros	3937.5	1329689212	2366139	998365	9708275
numPartnREC	Número de socios asignados al sector de investigación	0	56	1.01	0	2.29
numPartnHES	Número de socios asignados al sector de servicios de educación superior	0	131	1.67		3.01
NumPartners_EU13	Número de socios de países de Europa Central y del Este	0	53	0.42	0	1.38
NumPartners_EU15	Número de socios de países del grupo EU15	0	192	3.92	1	6.72
NumPartners_NonEU	Número de socios no europeos	0	42	0.59	0	1.51
acquaintance	Número de participaciones conjuntas de socios de proyectos en proyectos		30126	54.56		467.55
num_patentes	Número de patentes	0	80	0.07	0	0.86
share_unis	Porcentaje de universidades	0	1	0.48	0.4	0.44
share_resCen	Proporción de centros de investigación en el consorcio	0	1	0.18	0	0.31
share_compan	Proporción de compañías en el consorcio	0	1	0.29	0	0.39
share_EU13	Porcentaje de países EU13	0	1	0.06	0	0.19
share_EU15	Porcentaje de países EU15	0	1	0.81	1	0.33
share_nonEU	Porcentaje de países no europeos	0	1	0.13	0	0.29
NPub_peerArticle	Número de publicaciones por artículo	0	6033	5	0	35.91
NPub_ConfProceed	Número de publicaciones: conference	0	232	1.87	0	7.78
NPub_Other	Número de publicaciones de otros	0	1531	0.63	0	9.11
`NPub_non-peerArticle	Número de publicaciones: non-peer reviewed articles	0	91	0.11	0	1.06
NPub_ThesisDiss	Número de publicaciones de tesis	0	52	0.09	0	0.87
NPub_Books	Número de publicaciones de libros	0	52	0.06	0	0.59
NPub_BookChapt	Número de publicaciones de capítulos de libros	0	82	0.31	0	1.79
NPub_total	Número total de publicaciones	0	6196	8.08	1	40.17
NEntreg_Docs_reports	Número de entrega de documentos y reportes	0	183	2.45	0	6.74
NEntreg_Other	Número de entrega de otros	0	70	0.38	0	1.81
NEntreg_OpenResData	Número de entregables: Open Research Data	0	21	0.09	0	0.39
NEntreg_Websites_videos	Número de entregas de sitios web y videos	0	35	0.4	0	1.32
NEntreg_Demos_Prototyp	Número de entregables: Demos y prototipos	0	72	0.18	0	1.24
NEntreg_total	Número de entregas totales	0	207	3.51	0	8.77

Nombre de variable	Descripción	Mínimo	Máximo	Media	Mediana	Desviación estándar
ln_totalCost	Logaritmo de costo total	8.28	21.01	13.44	13.81	1.62
ln_ecMaxContribution	Log de contribución (\$) de la Comisión Europea	8.06	20.34	13.34	1.63	1.63
EC_cost_share	Proporción de la contribución (\$) de la Comisión Europea sobre el presupuesto total del proyecto	0.03	1	0.92	1	0.14
coord_exper_FP7	Experiencia del coordinador en coordinación de proyectos del FP7	0	664	69.02	18	125.55
coord_ranking_p1	Posición del coordinador en el ranking GWTS Leiden 2022	0	1204	34.68		111.71
particip_consorc_FP7	Experiencia de los miembros del consorcio de participación en proyectos del FP7	0	16340	352.54	115	628.01
coordin_consorc_FP7	Experiencia de los miembros del consorcio de coordinación en proyectos del FP7	0	8448	169.65	58	298.99

Fuente: elaboración propia con datos usados en el modelo

Fórmula de cálculo del índice de impacto:

El índice se calcula con base en variables escaladas y centradas. Se toma la versión logarítmica para disminuir el impacto de valores extremos. Se escalan las diferentes variables según ponderaciones de impacto.

$$\text{Índice}_{\text{impacto}} = \log (100 + \text{patentes} \times 10 + \text{Artículos}_{\text{pares}} \times 60 + \text{Publicaciones}_{\text{otras}} \times 10 + \text{Entregables}_{\text{otros}} \times 20)$$