

Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado

CARRERA DE ESPECIALIZACIÓN EN MÉTODOS
CUANTITATIVOS PARA LA GESTIÓN Y ANÁLISIS DE
DATOS EN ORGANIZACIONES

PROYECTO
TRABAJO FINAL INTEGRADOR DE ESPECIALIZACIÓN

Los determinantes de la inversión en Europa.
Análisis multivariado de la IED en 5 países europeos
entre 2010 y 2019.

AUTOR: GABRIEL OLIVER CHAMERO BANKS

MENTOR: ROBERTO ABALDE

MARZO 2021



Resumen

El objetivo de este trabajo es crear un modelo para predecir en qué sector se va a encuadrar una nueva inversión en los 5 países con el PIB más grande de la UE. Tras la salida definitiva del Reino Unido de la Unión Europea el 31 de diciembre de 2020, estos países son Francia, Alemania, Italia, España y Países Bajos.

Se tomará como referencia una base de datos de inversiones acometidas entre 2013 y 2018 clasificadas por sector CNAE (una codificación numérica que indica una actividad económica¹) elaborada a partir de los datos suministrados por la OCDE. Esta base de datos se completará con 19 determinantes de la IED (inversión extranjera directa) distribuidos en cuatro grandes áreas (Mercado, Talento e Innovación, Costes y Entorno de Negocios) elaborados a partir de información suministrada por la empresa Sifdi. Esta pequeña empresa española se especializa en el asesoramiento sobre inversiones extranjeras.

Para detectar patrones de comportamiento de los inversores se aplicará la técnica de Análisis de Componentes Principales. A continuación, se elaborará un modelo predictivo de clasificación para encuadrar la inversión en uno de los posibles sectores de actividad económica (Agricultura Ganadería y Pesca, Construcción, Servicios, Energía, Tratamiento de Agua, Industrias Extractivas). Por último, se redactará un informe con los resultados y las principales conclusiones del trabajo.

Se trabaja con la Hipótesis de que los determinantes de la inversión permanecen más o menos constantes a largo plazo: al analizar un periodo de tiempo relativamente largo (cinco años) en el que no ha habido crisis económicas de gran calado, se podrá extraer patrones en el comportamiento de los inversores.

Palabras clave

Atracción IED, determinantes, modelos predictivos, análisis componentes principales.

¹ Extraído de <https://www.codigoscnae.es/>



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Índice

Introducción	4
Gestión de datos en contextos organizacionales.....	7
Descripción metodológica.....	13
Implementación	26
Conclusiones	30
Bibliografía	32



Introducción

La IED ((Inversión Extranjera Directa, es decir, aquella inversión internacional que refleja el objetivo de una entidad residente en una economía de obtener una participación duradero en una empresa residente en otra economía)² forma parte de sistemas económicos internacionales abiertos y eficaces, siendo un catalizador importante del crecimiento económico³. Las previsiones económicas para el 2020 y 2021 apuntan a una dramática caída de los flujos de IED en las principales economías europeas. En este sentido, el *World Investment Report* de la UNCTAD prevé una caída del 40% en 2020 y una caída del 5% en 2021, la mayor caída en los flujos de inversión desde la Segunda Guerra Mundial⁴. En vista de la incertidumbre actual y los inmensos costes provocados por la crisis del coronavirus, los proyectos de inversión de muchas empresas se encuentran en suspenso.

En consecuencia, un modelo predictivo que detecte patrones de comportamiento de los inversores puede servir como guía a las autoridades económicas a la hora de diseñar sus planes para atraer IED una vez superada la pandemia, pues se presupone que los determinantes de la inversión que son específicos a cada sector permanecerán más o menos constantes. El año pasado, Europa pudo afianzarse como un lugar de inversión atractivo, con un total de 6.412 proyectos de inversión de empresas extranjeras. Esta cifra representó un uno por ciento más que en 2018 y el segundo nivel más alto jamás medido. Así lo demuestra la última encuesta realizada por la firma de auditoría y consultoría EY sobre proyectos de inversión de empresas extranjeras en Suiza y Europa⁵. En este contexto, es previsible que la competencia por atraer IED se intensifique dado el impacto económico de

² OECD. (1996). OECD benchmark definition of foreign direct investment. OECD.

³ Tenorio, M. A. L., & Sánchez, M. M. L. (2017). Impacto de la Inversión Extranjera Directa en el Desarrollo Económico-Social de la Región. *Jóvenes En La Ciencia*, 2(1), 697-701.

⁴ UNCTAD (2020), *World Investment Report 2020: International Production Beyond the Pandemic*, UN, New York, recuperado de <https://doi.org/10.18356/920f7642-en>.

⁵ EY (junio 2020). Massive decline in foreign direct investment expected in Switzerland and Europe, EY. recuperado de https://www.ey.com/en_ch/news/2020/06/massive-decline-in-foreign-direct-investment-expected-in-switzerland-and-europe



la pandemia de Covid-19. Todas las autoridades económicas (ayuntamientos, gobiernos, agencias de desarrollo económico, agencias de promoción de inversiones, etc.) tendrán que analizar cuáles son sus principales fortalezas y en qué tipo de sectores quieren atraer inversión. Por todo ello cobra especial relevancia la siguiente pregunta: ¿Cuáles son los determinantes de la inversión que tienen más importancia a la hora de encuadrar una inversión en un sector determinado?

El objetivo principal de este trabajo es elaborar un modelo que prediga en qué sector se va a encuadrar una nueva inversión europea, en base a ciertas características (los determinantes de la inversión) que suelen atraer IED. El objetivo secundario del trabajo es identificar los determinantes de la inversión que más influyen en la toma de decisiones de los inversores.

La Hipótesis del trabajo es que los determinantes de la inversión permanecen más o menos constantes a largo plazo. Al analizar un periodo de tiempo relativamente largo en Europa (cinco años), en el que no ha habido crisis económicas de gran calado, se podrán extraer patrones en el comportamiento de los inversores. En este sentido asumiremos como válidas las tesis de Dornean y Oanea⁶. Según estos autores:

“todos los tipos de determinantes (infraestructura, capital humano y estabilidad económica) tienen un impacto significativo en la IED. El atractivo de inversión de un país depende del comportamiento de inversiones anteriores, por lo que es importante para el país anfitrión aplicar políticas sostenibles que promuevan la IED a largo plazo y no solo a corto plazo”.

Quizás estos patrones de comportamiento puedan servir como guía a las autoridades económicas a la hora de diseñar sus planes para atraer IED una vez superada la pandemia.

Se comenzará por armar la base de datos primaria, relacionando las inversiones acometidas durante los últimos 5 años en los países citados anteriormente por sector (con datos de la OCDE) con los principales determinantes de la inversión. A continuación, se hará un análisis de componentes principales con el lenguaje de programación R con la finalidad de

⁶ Dornean, A., & Oanea, D. C. (2015). Impact of the economic crisis on FDI in Central and Eastern Europe. *Review of Economic and Business Studies*, 8(2), 53-68.



detectar patrones en el data set. Habiendo estudiado las variables mediante estadística descriptiva, se elaborará un modelo predictivo usando la plataforma Azure Machine Learning Studio de Microsoft. Por último, se elaborará informe con los principales resultados y conclusiones del análisis. El siguiente diagrama de Gantt muestra las distintas etapas en la consecución del proyecto:

Actividad	Temporalidad								
	1	2	3	4	5	6	7	8	9
Diseño y limpieza del data set	■	■	■	■					
PCA con R					■	■			
Elaboración del modelo predictivo con Azure ML Studio							■	■	
Elaboración del Informe								■	■

Tabla 1 – Diagrama de Gantt



Gestión de datos en contextos organizacionales

Descripción de la organización

Sifdi es una empresa privada independiente que brinda soluciones eficientes para proyectos de inversión extranjera directa (IED) en España y en aquellos países donde opera su red internacional. Ofrece asesoramiento confidencial a empresas internacionales que planean inversiones por primera vez y a aquellas que buscan expandir o diversificar sus actividades en esas regiones. Sifdi también elabora una *newsletter* mensual con las principales noticias sobre la inversión extranjera en España. La empresa tiene una amplia experiencia en España, Portugal y más de 10 países en América Latina⁷.

La gran mayoría de los clientes de Sifdi son agencias públicas o agencias de promoción de inversiones a las que apoya en la elaboración o revisión de la propuesta de valor de su localización. En estos informes se describe la situación de la ciudad, región o país objeto del análisis y se resaltan las características diferenciales del territorio frente a otros potenciales destinos, ayudando así en la creación y consolidación de la imagen de marca deseada. A estas agencias también se les proporciona información sobre *leads*, es decir, potenciales inversores que pueden tener un interés en invertir en su territorio.

En cuanto a su composición, Sifdi es el resultado del esfuerzo común de un grupo de experimentados profesionales especializados en los ámbitos de comercio exterior, la internacionalización y la atracción de inversiones. Debido a su reducido tamaño (menos de diez empleados), muchos roles son asumidos por una misma persona. El CEO de la empresa, por ejemplo, también actúa como director de TIC, mientras que la directora para Latinoamérica ejerce también como directora de comunicaciones.

⁷ Extraído de <http://sifdi.com/nosotros.html>



En sus inicios, el modelo de negocio de Sifdi se planteó como lineal: la empresa toma componentes/insumos (datos de registros de inversiones, mayoritariamente), crea servicios terminados (informes estratégicos) y los vende a sus consumidores (agencias de promoción de inversiones o empresas privadas). El valor que aporta la empresa está en los activos / valor añadido que ofrece en sus informes. Su modelo es el de una consultora tradicional, es decir, se contrata a los mejores talentos y se cobra a los clientes ya sea una tarifa fija por hora o día para obtener acceso a este talento y a su experiencia, o una tarifa negociada por cada informe.

No obstante, el modelo de negocio de Sifdi ha ido evolucionando y en la actualidad Sifdi colabora también con dos empresas con amplia experiencia en negocios y comercio internacional, CFI⁸, una empresa que brinda soluciones financieras para proyectos de exportación e inversión e Iberglobal⁹, una consultora de comercio exterior especializada en la realización de estudios sobre internacionalización de la economía y de la empresa. Esta colaboración se basa en un modelo de *collaborative consulting*¹⁰, en el que el principal objetivo de la colaboración es aportar un mayor valor a los clientes de la empresa.

Gestión de datos por parte de la organización

En cuanto a la gestión de datos, si nos atenemos a la clasificación del *MIT Sloan Review*¹¹, Sifdi está transformándose para pasar de ser *Aspirational* (focalizada en mejorar procesos y costos y con pocos recursos disponibles para DA (*Data Analytics*), a ser *Experienced* (ir más allá del análisis coste/beneficio y buscar otros resultados más significativos, nuevas percepciones o *insights*). Actualmente, la empresa no cuenta con un departamento especializado en Data Management y muchas veces subcontrata servicios como el procesamiento (generación de visualizaciones, por ejemplo) a empresas externas como Graphext. Para almacenar la información (mayoritariamente en forma de informes y

⁸ Extraído de <https://cfispain.com/en/quienes-somos/>

⁹ Extraído de <https://iberglobal.com/index.php/quienes-somos>

¹⁰ Consultancy UK, December 2019, extraído de <https://www.consultancy.uk/news/23120/three-new-business-models-in-the-consulting-industry>

¹¹ LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT sloan management review*, 52(2), 21-32.



publicaciones), Sifdi usa Notion, un repositorio virtual. Para el almacenamiento de los datos ‘procesados’ de cada una de las fuentes se usa Excel.

No obstante, a medida que Sifdi incorpore cada vez más datos a su operativa, deberá adoptar una arquitectura de datos moderna, y el talento adecuado para gestionarla (especialistas en ciencia de datos, arquitectos de datos, etc.). Para ello, se propone una arquitectura que sea capaz de capturar los datos estructurados y no estructurados, procesarlos (crear tablas y archivos de texto) y publicar un reporte / *newsletter*, con las visualizaciones pertinentes. El esquema planteado sería el siguiente:

- Azure Data Factory para la captura de los datos estructurados y no estructurados
- Azure Data Lake para almacenamiento de datos estructurados y no estructurados
- Polybase para pasar de CSV a tablas (BBDD de SQL) / Parseur para traspasar datos de las alertas a archivos de texto
- Power BI para generar los informes y visualizar la información obtenida

En Sifdi, el proceso de toma de decisiones depende del servicio contratado y de si el cliente es un organismo público o una empresa privada. Si el cliente es del sector público, suele haber una fase inicial de promoción en la que se le informa al cliente de los servicios disponibles. Cuando el organismo público consigue presupuesto, se abre un proceso de presentación de propuestas. Si la empresa (Sifdi) lo estima oportuno, presenta una oferta y en caso de resultar ganadora, ejecuta su propuesta. Con empresas privadas el proceso es similar. La empresa contacta con Sifdi, se elabora una propuesta preliminar ad-hoc en la que se estima la cantidad de recursos y los retornos, etc. y se presenta una propuesta definitiva. En este sentido, los datos que Sifdi usa para la toma de decisiones son puramente económicos, es decir, se evalúa el proyecto en cuestión desde el punto de vista de los costes que supone la operación y el retorno esperado.



Al igual que en la mayoría de las consultoras especializadas del sector, las decisiones las toma el HiPPO (*Highest Paid Person's Opinion*)¹², o persona mejor pagada dentro de la empresa, basadas en su conocimiento del mercado e intuición). En el caso de Sifdi, el HiPPO es el CEO de la empresa. No obstante, en Sifdi, al igual que en muchas otras empresas especializadas en la atracción de inversiones, cada vez está adquiriendo más importancia la toma de decisiones basadas en datos (*data driven decision making*)¹³. Estas empresas están mejorando su estrategia tecnológica con plataformas de sistemas de información geográfica, aplicaciones *proptech* (empresas que están adoptando nuevas herramientas a la compraventa de casas o a cualquier otro apartado relacionado con el inmobiliario)¹⁴ e incluso realidad virtual para atender a empresas en sus procesos de selección de ubicaciones. Por ejemplo, las soluciones basadas en la tecnología de sistemas de información geográfica (GIS) está ayudando a las autoridades y empresas locales a combatir las ineficiencias típicas en el segmento comercial del mercado inmobiliario.

Problemática de la organización y la gestión de los datos

Al ser una empresa pequeña, Sifdi no se ve tan afectada por el problema de que haya inconsistencias en sus datos maestros (datos de clientes, empleados, proveedores, etc.), datos de referencia (esquemas de clasificación a los que se hace referencia para almacenar datos, ejecutar procesos y generar informes, como listas de valores válidos, listas de códigos, códigos de estado, abreviaturas de estado, campos demográficos, etc.) o transaccionales (aquellos que describen un evento o transacción interna o externa que tiene lugar mientras una organización realiza su negocio, como órdenes de venta, facturas, órdenes de compra) como podría estarlo una empresa multinacional. Por el contrario, uno de los principales desafíos en término de datos a los que se enfrenta Sifdi está el mantener una base de datos actualizada con las principales inversiones acometidas en España / Madrid, o, en su defecto, el territorio concreto que solicite los servicios de la empresa. Pare

¹² Deb Gallagher (2012). MIT Sloan Review, Extraído de <https://sloanreview.mit.edu/article/the-decline-of-the-hppo-highest-paid-persons-opinion/>

¹³ Brynjolfsson, E., & McElheran, K. (2016). The rapid adoption of data-driven decision-making. *American Economic Review*, 106(5), 133-39.

¹⁴ Extraído de https://www.businessinsider.es/que-es-proptech-tecnologia-mercado-inmobiliario-189802?gclid=Cj0KCQiA88X_BRDUARIsACVMYD8R28VmFCoMH07YvSuRTEyyv4QcR2uWtIukbxswU0eAqyyEklb3hO8aAq1AEALw_wcB



ello, elabora un registro de las principales inversiones acometidas en España durante un periodo de tiempo determinado, o, en su defecto, el territorio concreto que solicite los servicios de la empresa. Así, por ejemplo, saber cuánta inversión extranjera (número de proyectos totales) ha entrado en España en un periodo de tiempo determinado es uno de los grandes retos de la compañía en términos de datos.

Las fuentes de información secundaria que usa Sifdi son, en su mayoría, datos abiertos de organismos públicos. Entre ello podemos citar la UNCTAD (el *World Investment Report* y su base de datos con estadísticas de inversiones), los informes y estadísticas de organismos como la OCDE y Eurostat, así como la información suministrada por el Instituto Nacional de Estadística de España y el Registro de Inversiones en España. Todos estos registros son datos estructurados (se presentan con formato tabular) y de periodicidad *batch* (trimestral en el caso de los datos del Registro de Inversiones/INE y anual en el caso de la OCDE, Eurostat y del INE-FILINT (Estadística de Filiales de Empresas Extranjeras). También se usan informes de consultoras privadas como KPMG, Cushman & Wakefield, Deloitte, CBRE, etc. que se suelen publicar anualmente (*Office Sublease Market Report* de Cushman & Wakefield, el *Global Office Rent Tracker* de CBRE, etc.).

La principal fuente de información primaria que maneja Sifdi es las entrevistas que hacen a altos mandos del mundo empresarial en las que describen su experiencia a la hora de invertir en España. Estas entrevistas sirven como base para asesorar a los clientes de la empresa. Además de estas entrevistas a directivos, también se hacen entrevistas a responsables de agencias de promoción de inversiones, expertos en el ámbito de la Inversión y a diversos multiplicadores/prescriptores. Para la elaboración de su *newsletter*, Sifdi usa algoritmos/filtros como los de Google (<https://www.google.es/alerts>) sobre colectivos acotados. Mediante las herramientas de Graphext/Contexto se definen varios grupos de corresponsales extranjeros en España y de personas relacionadas con la inversión extranjera directa (IED), las cuales envían alertas si estos colectivos usan determinadas palabras.



En cuanto a las problemáticas asociadas con el uso de la información disponible cabe destacar el uso dado a las entrevistas y a la confidencialidad de los datos. Los riesgos asociados a estos datos son principalmente dos: que sean reutilizados, es decir, que los datos originalmente recopilados para un propósito específico y se usen de nuevo para fines comparables en dominios comparables o que los datos suministrados sean reasignados: (tomar los datos originalmente recopilados para un propósito específico en un dominio específico y analizarlos para fines no relacionados en un dominio que no sea su dominio de origen). Para mitigar estos riesgos, Sifdi pide permiso expreso a la hora de publicar la entrevista y usar la información suministrada. En este sentido, Sifdi cumple la normativa europea del Reglamento General de Protección de Datos (RGPD) y la Ley Orgánica 3/2018, de 5 de diciembre, de Protección de Datos Personales y garantía de los derechos digitales vigente en España. Además, para los datos que Sifdi usa en proyectos concretos (planes estratégicos, planes de negocio, etc.) se compromete a no usar los datos que le faciliten desde agencias/gobiernos de otros proyectos, salvo consentimiento explícito suyo. Del mismo modo, Sifdi pide a sus clientes que aseguren la confidencialidad de los datos a los que puedan tener acceso.

A medida que Sifdi incorpore a su modelo de negocio tecnologías de Big Data, como el uso de imágenes de geolocalización anteriormente citado, deberá tener en cuenta cómo afecta el uso de estas tecnologías a la privacidad. Respecto a los datos que provienen de las fuentes usadas para los registros de inversiones (fuentes secundarias), uno de los principales problemas es el tratamiento de los datos faltantes, pues en muchos casos la información para un año / localidad en concreto no está disponible. Otro problema relacionado con la operativa de los datos es el derivado de las distintas definiciones de lo que constituye inversión extranjera por parte de cada fuente, y la metodología que cada una usa para medirla. Esto puede dar lugar a inconsistencias en los informes que se publican. Asimismo, los retrasos en la publicación de los datos por parte de algunas fuentes (por ejemplo, los datos del INE con los que trabaja actualmente la empresa son de 2018) implica que los informes publicados no necesariamente reflejen la realidad del momento. Por último, cabe destacar cierto desconocimiento en general de España por fuentes



internacionales, anglosajonas en su gran mayoría, lo cual implica cierta falta de representatividad de España en sus estadísticas.

Descripción metodológica

Recopilación de la información

El set de datos usados en este trabajo proviene de la sección de datos abiertos de la OCDE https://stats.oecd.org/Index.aspx?DatasetCode=FDI_FLOW_INDUSTRY y se compone de 10468 instancias de inversiones acometidas en las cinco economías más grandes de la UE (Alemania, Francia, Italia, España y Países Bajos) entre los años 2010 y 2019. La variable para predecir (el *label*) es el Sector Económico de la inversión (*Industry*), de acuerdo con la clasificación de la OCDE establecida en la tercera edición de su *Benchmark Definition*: (<http://www.oecd.org/investment/investmentpolicy/2090148.pdf>). Según esta clasificación, los posibles destinos de la inversión son: Sector Primario (Agricultura y Pesca), Minería y Canteras, Manufacturas, Electricidad, Gas y Agua, Construcción y Servicios. El número de operaciones por sector económico de actividad se muestra en la siguiente tabla:

:

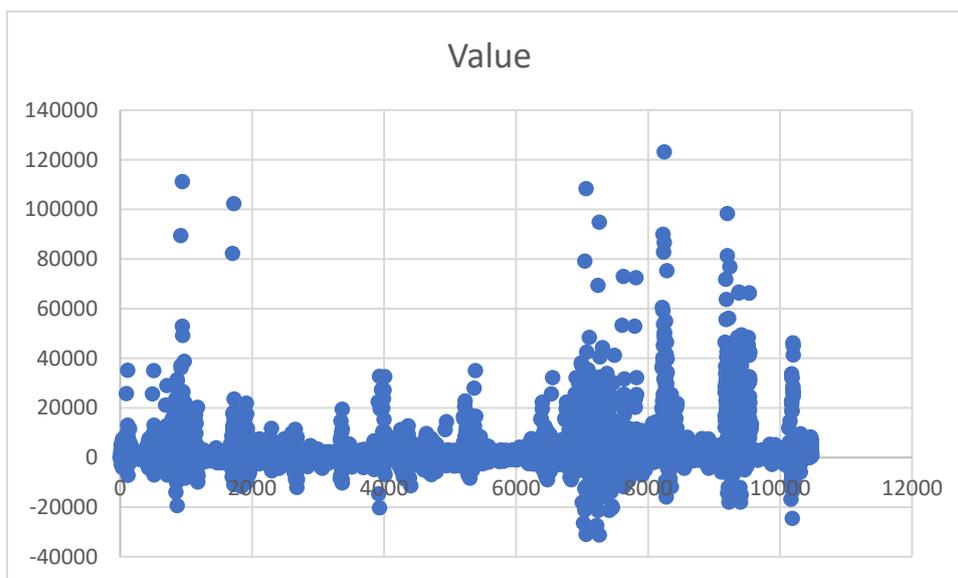
Industry	Number of operations per Industry (2013-2018)
CONSTRUCTION	200
ELECTRICITY, GAS AND WATER	196
MANUFACTURING	2840
MINING AND QUARRYING	584
PRIMARY SECTOR	396
SERVICES	6066
Unallocated	200

Tabla 2 – Volumen de operaciones por Sector de Actividad



Variables Predictoras

Value (valor de la inversión). Se considera IED aquella inversión en la que el objetivo del inversor es obtener un interés duradero (propiedad del 10 por ciento de las acciones ordinarias o con derecho a voto) en una entidad residente en una economía distinta a la suya. El valor medio de la inversión para la serie objeto de nuestro análisis es de 27.5 M de euros. Los valores que puede tomar esta variable se reflejan en la siguiente gráfica:



Gráfica 1: Valores de la inversión

Reporting Country: el país que recibe la inversión

Francia (2184 inversiones), Alemania (2192), Italia (1968), España (1962) y Países Bajos (2176).

Doing Business: este indicador es una publicación insignia del Grupo Banco Mundial. Se trata de una serie de estudios anuales que evalúan las regulaciones que favorecen o restringen la actividad empresarial. El indicador se basa en la *Distance to Frontier*, es decir



la distancia de cada economía a la "frontera", la cual representa el mejor desempeño observado en cada uno de los indicadores en todas las economías de la muestra de *Doing Business* desde 2005. A mayor puntuación, más facilidad para hacer negocios en el país.

PIB per cápita: un indicador económico que mide la relación existente entre el nivel de renta de un país y su población. Para ello, se divide el Producto Interior Bruto (PIB) de dicho territorio entre el número de habitantes. Los datos corresponden al valor medio registrado para los cinco años en los que se basa el estudio y provienen de la OCDE.

Stock de IED recibida: indica el valor acumulado medio como porcentaje del PIB de todos los tipos de inversión para los cinco años objeto del análisis según datos de la UNCTAD (*United Nations Conference on Trade and Development*). El valor de Países Bajos (104) es excepcionalmente alto debido a que muchas inversiones se canalizan a través de este país debido a su atractivo sistema fiscal.¹⁵

Labor costs – Este indicador indica el coste medio por hora en Euros de un trabajador de acuerdo con las estadísticas de Eurostat. En 2019, los costes laborales medios por hora se estimaron en 27,7 EUR en la UE-27 y en 31,4 EUR en la zona del euro (EA-19). Sin embargo, este promedio enmascara brechas considerables entre los Estados miembros de la UE, con costes de mano de obra por hora que oscilan entre 6,0 EUR y 44,7 EUR. El dato utilizado para cada país corresponde a la media de los cinco años objeto del estudio.

Logistics performance Index: un indicador del Banco Mundial que refleja la capacidad logística del país. El LPI es una herramienta interactiva de evaluación comparativa creada para ayudar a los países a identificar los desafíos y oportunidades que enfrentan en su desempeño en logística comercial y lo que pueden hacer para mejorar su desempeño.

Velocidad Internet – Datos de la Comisión Europea que reflejan la velocidad media de bajada de datos para los cinco años objetos del estudio. Los datos provienen del Índice de Economía y Sociedad Digitales (DESI), un índice compuesto que resume los indicadores

¹⁵ <https://www.taxjustice.net/2014/11/10/netherlands-worlds-largest-source-fdi/>



relevantes sobre el desempeño digital de Europa y rastrea la evolución de los Estados miembros de la UE en competitividad digital.¹⁶

Coste elect ind – Coste medio a lo largo de los cinco años objetos del estudio de la electricidad para la industria en Megavatios/hora según datos de Eurostat. Los usuarios finales industriales pueden incluir datos de otras áreas como oficinas, restaurantes, servicios, transporte, etc.

Nivel de Inglés – Puntuación del país en el ranking elaborado por la consultora EF (*Education First*), especializada en la enseñanza de inglés como idioma secundario. El índice de EF evalúa el dominio del idioma inglés en los adultos y clasifica los países y regiones no angloparlantes según los resultados obtenidos. Este informe se ha convertido en un índice de referencia para la mayoría de los gobiernos de todo el mundo y se elabora con los datos extraídos de la prueba de inglés EF SET.

Desarrollo Merc Fin – Desarrollo del mercado financiero. Este indicador publicado por el *World Economic Forum* proporciona una puntuación y clasificación para la amplitud, profundidad y eficiencia de los principales sistemas financieros y mercados de capitales del mundo. El Índice analiza los impulsores del desarrollo del sistema financiero que respaldan el crecimiento económico y, por lo tanto, compara la competitividad general de los sistemas financieros. El valor reflejado para cada país corresponde a la media para los cinco años objetos del estudio.

Índice de Innovación – puntuación media para los cinco años objeto del análisis en el *Global Entrepreneurship Monitor*. El GEM, a través de una metodología común de medición en todos los países involucrados, describe al emprendedor y su entorno y ayuda a identificar los factores que incrementan los niveles de emprendimiento. Los resultados constituyen una valiosa herramienta de comparación entre países y son utilizados como referente por instituciones que diseñan e implementan procesos de apoyo al emprendimiento en todo el mundo.

¹⁶ Extraído de <https://ec.europa.eu/digital-single-market/en/digital-economy-and-society-index-desi>



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



Country RepTrack – puntuación el ranking del *Reputation Institute*. Esta empresa publica informes sobre la reputación de corporaciones, ciudades y países basados en encuestas a consumidores y cobertura de medios. El ranking *Country RepTrak* clasifica la reputación de los países analizados utilizando tres criterios: "entorno atractivo", "economía avanzada" y "gobierno eficaz".

Impuestos – indicador que refleja el porcentaje que se aplica al impuesto de sociedades (según datos de la consultora KPMG). Esta tabla de impuestos corporativos proporciona una vista de las tasas de impuestos corporativos en todo el mundo. El número refleja el impuesto medio de cada país para los cinco años objeto del estudio

Calidad de Vida – indicador que refleja el equilibrio entre vida personal y laboral elaborado a partir del Better Life Index de la OCDE. El Índice para una Vida Mejor se creó para permitirte visualizar y comparar algunos de los factores clave —como educación, vivienda, medio ambiente y otros— que contribuyen al bienestar en los países de la OCDE. El índice permite ver cómo se comportan los países de acuerdo con la importancia que se le da a cada uno de los 11 factores que hacen posible una vida mejor.

Producción científica – refleja el número de documentos elaborados según *SCImago Journal Rank*, un ranking de revistas, países e instituciones elaborado por un grupo de investigación interuniversitario español (universidades de Granada, Extremadura, Carlos III de Madrid y Alcalá de Henares). Su principal indicador, el Índice SJR, analiza las citas recibidas por las revistas con una ventana de tres años y pondera las citas dando más valor a las citas provenientes de revistas con mayor índice SJR.



Procesamiento de la información

Exploratory Data Analysis

Ejecutamos el comando *Summarize Data* en Azure Machine Learning Studio. Los resultados los podemos ver en la siguiente tabla

Feature	Count	Unique Value Count	Valores Faltantes	Min	Max	Mean	Mean Deviation
Industry	10482	7	0				
Stock IED	10482	5	0	18.7	104	44.10013	25.281444
Doing Business	10482	5	0	72.85	79.71	76.73499	1.72246
PIB per capita	10482	5	0	36318	50540	43348.02	5327.747566
LPI	10482	5	0	3.74	4.2	3.932003	0.148622
Velocidad Internet	10482	5	0	17.3	40.21	29.8402	6.883787
Coste Elect.	10482	5	0	0.1044	0.1967	0.142114	0.031124
Desarrollo Merc Fin	10482	5	0	3.5	5	4.34398	0.445686
Nivel de Inglés	10482	5	0	55.49	70.31	60.41174	5.501652
Indice Innovacion	10482	5	0	46.3	63.3	55.75105	4.111879
Country Reputation	10482	5	0	68.5	76.7	72.45035	2.964996
Impuestos	10482	4	0	24	30	26.48292	2.103173
Calidad de Vida	10482	5	0	7.5	9.3	8.576169	0.519402
Produccion Cientifica	10482	5	0	51285	153271	98117.97	25923.00514
Labour Costs	10482	4	0	15.9	26.4	22.84974	3.521558
Reporting country	10482	5	0				
Value	10482	6885	0	0	123192.3	2755.361	3265.622764

Tabla 3 – Exploratory Data Analysis

De acuerdo a este análisis, podemos establecer que:

- Hay 10482 filas o registros en nuestro dataset.
- El dataset tiene 17 variables o columnas.
- Las columnas tienen tipos de datos o bien chr (texto), num (numérico o número de punto flotante) o int (entero).

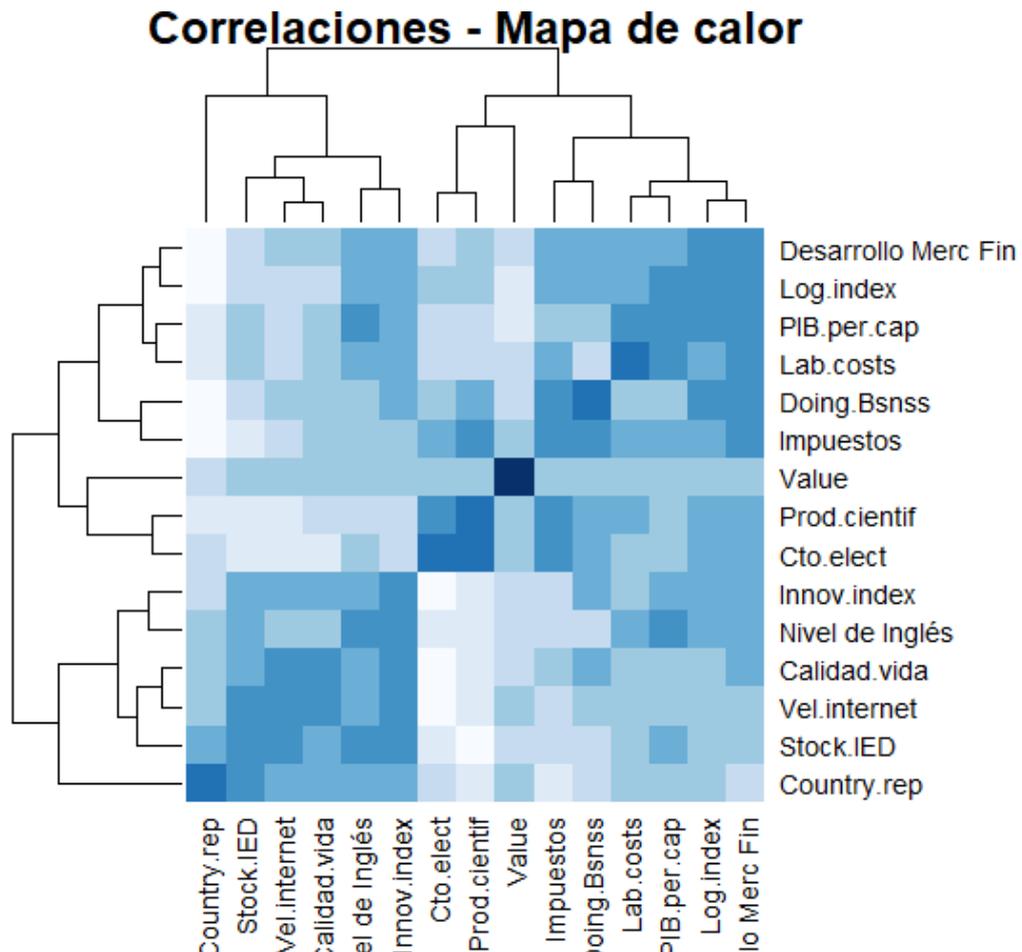
Correlación de las variables

Una de las mejores maneras para visualizar correlaciones entre variables es un mapa de calor (técnica de visualización de datos que muestra la magnitud de un fenómeno como color en dos dimensiones). La variación en el color puede ser por matiz o intensidad, dando señales visuales obvias al lector sobre cómo el fenómeno se agrupa o varía en el espacio. En este mapa de calor / dendrograma, el azul más intenso se asocia con una



correlación positiva entre variables, mientras que el color blanco es señal de una correlación negativa.

Como se puede observar, el dendograma permite diferenciar 2 grandes grupos de variables:



Gráfica 2 – Mapa de Calor / Dendograma

Preparación de los Datos (DataPrep)

Valores faltantes. La variable *Industry* tiene 20 valores faltantes. Dado que representan un porcentaje ínfimo del total de los datos, se opta por suprimir estas instancias. A su vez, la variable *value* (valor de la inversión) incluye 330 valores faltantes, todos ellos en el sector servicios. A estos valores se les imputa la media para este sector económico. Estas imputaciones se hacen directamente sobre el dataset en Excel.

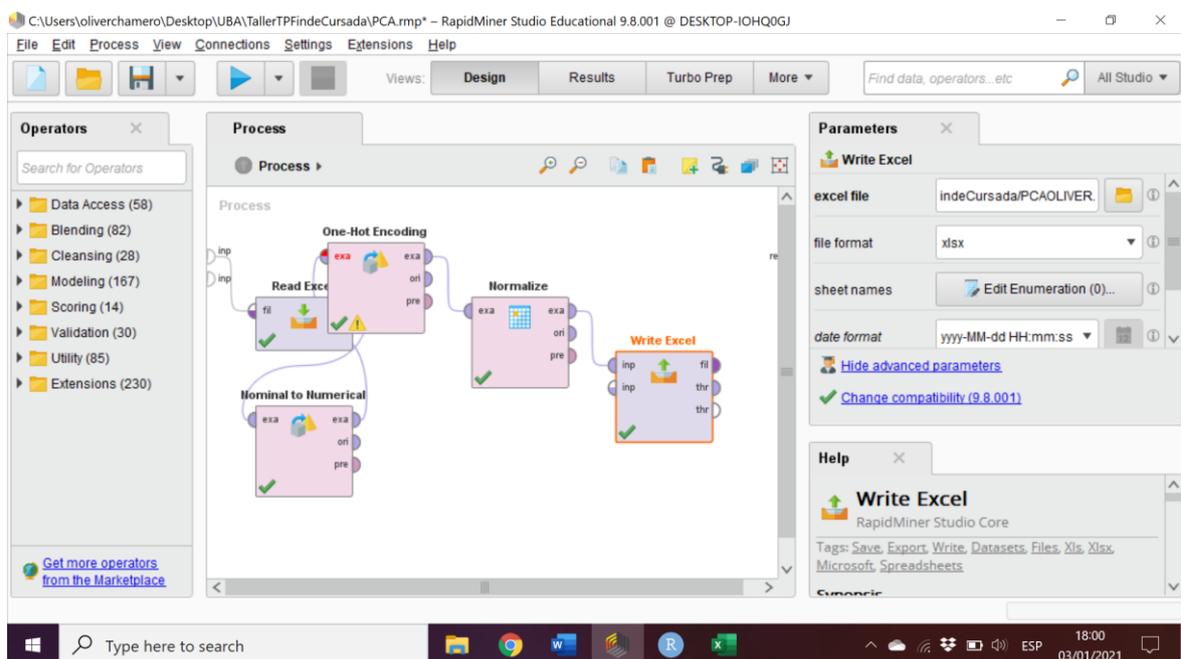


Agregación – Para facilitar el análisis se crean los cinco metagrupos mencionados anteriormente que engloban todos los sectores contenidos en el dataset original de la OCDE.

Convert to Indicator Values: Usamos la técnica de *One Hot Encoding* para transformar las variables ordinales en nominales (unos y ceros) (Reporting Country y Industry)

Split Data: Hacemos un *Stratified Split* para asegurarnos de que tenemos la misma proporción de la variable servicios dentro del set de entrenamiento y el data set de testeo. La variable servicios representa por si sola cerca del 60% de las inversiones

Para buscar patrones y reducir la dimensionalidad del dataset empleamos la técnica de Análisis de Componentes Principales (PCA). Previamente normalizamos el data set



Gráfica 3 – Preparación de datos para la ejecución del PCA



Análisis de Componentes Principales

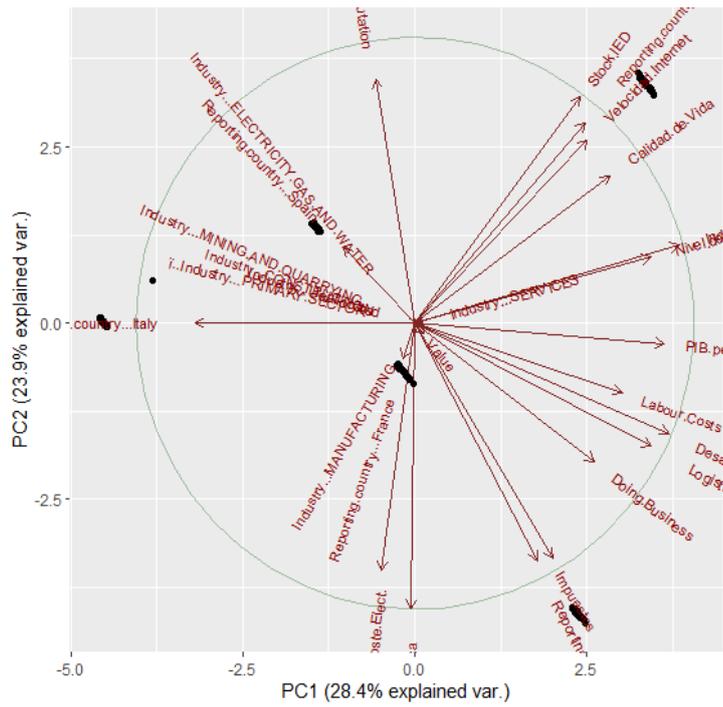
Al aplicar la técnica de Análisis de Componentes Principales logramos explicar cerca del 100% de la varianza a partir de los primeros diez componentes principales

summary(datos.prc)

Importance of components:

	PC1	PC2	PC3	PC4	PC5
Standard deviation	2.7668	2.5415	1.7313	1.38031	1.33145
Proportion of Variance	0.2835	0.2392	0.1110	0.07057	0.06566
Cumulative Proportion	0.2835	0.5228	0.6338	0.70434	0.77000
	PC6	PC7	PC8	PC9	PC10
Standard deviation	1.06546	1.02546	1.01650	1.01257	1.00954
Proportion of Variance	0.04204	0.03895	0.03827	0.03797	0.03775
Cumulative Proportion	0.81204	0.85099	0.88926	0.92724	0.96498

Parece que alrededor del 80% de la varianza acumulada se puede explicar en los primeros seis componentes (PC1 – PC6). Cerca del 90% de la varianza se puede explicar en los primeros nueve componentes (PC1 – PC9). La importancia de cada componente puede apreciarse en el siguiente gráfico:

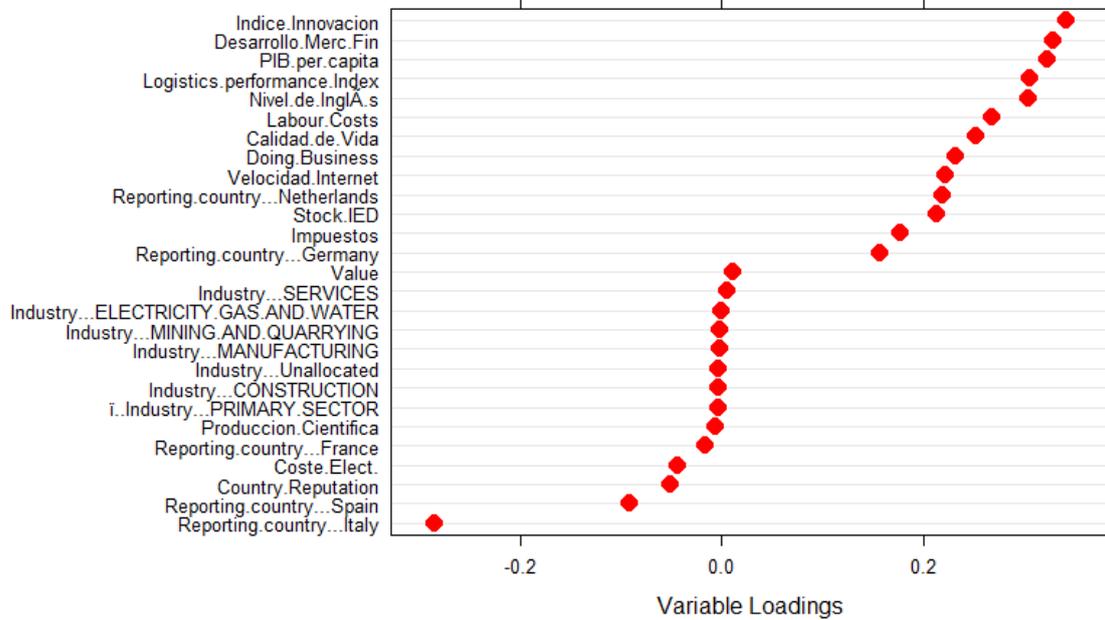


Gráfica 4 – GGPlot de las Componentes Principales

Como se puede apreciar, parece que Índice de innovación, Desarrollo de Mercados Financieros y PIB per cápita son las variables que más incidencia tienen en los *Loadings* del Componente Principal 1



Loadings Plot for PC1

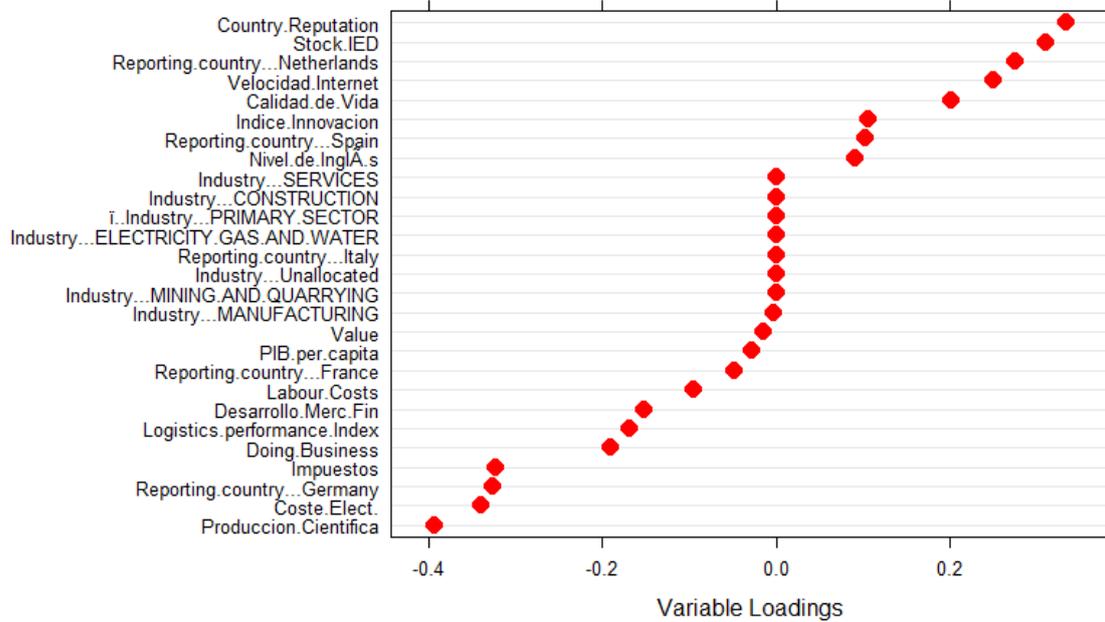


Gráfica 5 – Loadings PC1

Asimismo, de la siguiente gráfica se desprende que Country Reputation, Stock de IED y Reporting Country - Netherlands (Países Bajos) son las variables que más repercusión tienen en los *Loadings* del Componente Principal 2



Loadings Plot for PC2



Gráfica 6 - Loadings PC2

Análisis de la información

Se ejecutan varios algoritmos de clasificación (*Multiclass Decision Forest*, *Multiclass Logistic Regression*, *Multiclass Neural Networks* y *Multiclass Decision Jungle*). La performance del modelo con los mejores resultados (*Multiclass Decision Forest*) se puede observar en la siguiente tabla

Metrics	
Overall accuracy	0.477481
Average accuracy	0.850709
Micro-averaged precision	0.477481
Macro-averaged precision	0.177808
Micro-averaged recall	0.477481
Macro-averaged recall	0.174197

Tabla 3 – Performance del modelo.

Como podemos observar, la *average accuracy* (suma de precisión para cada clase predicha / número de clases) es muy superior a la *overall accuracy* (número de elementos predichos correctamente / total de elementos a predecir). Esto puede dar lugar a que el modelo no sea



muy fiable a la hora de predecir una clase específica, dejándose guiar por la clase que está prediciendo con la mejor proporción, aumentando así las predicciones incorrectas. Una visión más exacta del funcionamiento del modelo la podemos ver en la siguiente matriz de confusión. El modelo tiene un rendimiento más o menos aceptable al predecir el sector servicios, con una precisión de 63% de instancias acertadas. No obstante, el rendimiento con las demás clases es relativamente pobre (34% para Manufacturas, mientras que para el resto de las clases no se supera el 10% de aciertos)

Class	CONSTRUCTION	ELECTRICITY	GAS AND WATER	MANUF.	MINING AND	PRIMARY SEC	SERVICES	Unallocated	Average Log L	Precision
CONSTRUCTION	0	1	17	5	2	25	0	27.79323	0	0
ELECT, GAS AND WATER	0	1	20	2	0	25	1	24.4452521	0.02325581	0.02040816
MANUFACTURING	17	11	269	43	18	346	6	10.4780592	0.34137056	0.37887324
MINING	4	3	52	14	12	59	2	20.8750965	0.09459459	0.09589041
PRIMARY SECTOR	0	1	39	12	3	44	0	20.9480991	0.03658537	0.03030303
SERVICES	19	25	380	68	46	961	17	2.79924529	0.6453996	0.63390501
Unallocated	1	1	11	4	1	29	3	24.8373536	0.10344828	0.06

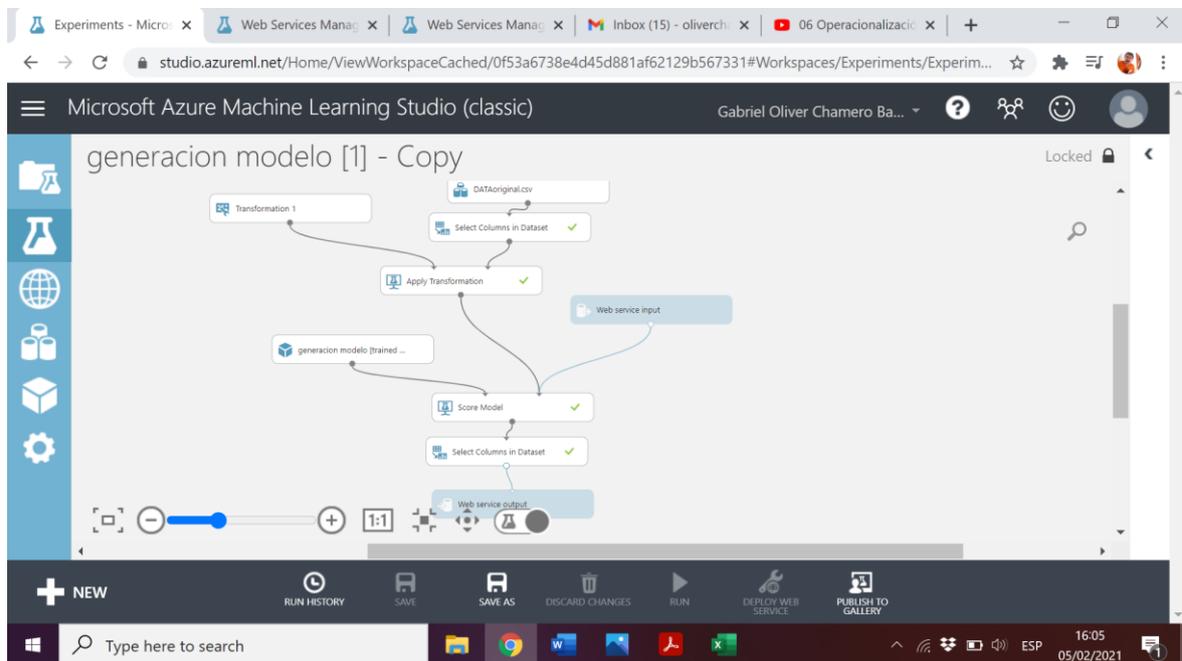
Tabla 4 - Matriz de confusión



Implementación

Operacionalización del modelo

Los pasos para la operacionalización del modelo se pueden ver en la toma de pantalla que se muestra a continuación



Gráfica 7 – Operacionalización del modelo

Como se puede observar en la siguiente toma de pantalla, el modelo no genera errores al ejecutarse, es decir, se ha puesto en producción correctamente.



The screenshot displays the Microsoft Azure Machine Learning Studio (classic) Web Services interface. The browser address bar shows the URL: `services.azureml.net/workspaces/0f53a6738e4d45d881af62129b567331/webservices/f2cc5bc3530e436e91fdb60b42125098/...`. The interface includes a navigation menu with options: Quickstart, Dashboard, Batch Request Log, Configure, Consume, Test (selected), and Swagger API. The main area contains a form with the following fields and values:

Field	Value
Impuestos	2
Calidad de Vida	6
Produccion Cientifica	23000
Labour Costs	23
Value	21000

Below the form is a green button labeled "Test Request-Response". Underneath, there is a section for "output1" which is currently empty. At the bottom of the interface, there are labels for "Scored Labels" and "SERVICES". The Windows taskbar at the bottom shows the search bar, system tray, and the date/time: 16:32 on 05/02/2021.

Gráfica 8 – Resultados del modelo

Dado que la cantidad de datos que se maneja en un principio es relativamente pequeña, se elige la configuración básica del hardware (instalada por defecto) al publicar el modelo en Azure ML. Lo más probable es que un único usuario sea el responsable de ejecutar el modelo. Debido a que se dispone de poco tiempo para clasificar las inversiones (el modelo se ejecuta en tiempo real al estar publicado el modelo como servicio online) y que el tiempo que tarda el modelo en ejecutarse no es un factor determinante de éxito) no se necesitará una infraestructura demasiado potente en términos de CPU o de memoria. No obstante, a medida que se incorporen nuevos predictores y el modelo tenga que manejar un mayor volumen de datos, es posible que se requiera una configuración más potente.

Monitoreo

El control sobre lo que sucede al ejecutar el modelo se hace en el log. El logueo permite comprobar si las predicciones hechas por el modelo son fiables. Los logs del modelo se guardarán por 90 días, una ventana de tiempo suficiente para comprobar si el modelo funciona correctamente. El almacenamiento de los logs se efectuará sobre la plataforma de Azure Storage – no se prevé usar una cantidad muy grande de espacio, pues, como dijimos anteriormente, el volumen de consultas sea bajo al haber un único usuario de modelo. Sin



embargo, el modelo se irá actualizando con nuevas versiones a medida que se vayan incluyendo nuevos predictores.

Calidad de las predicciones

Al ser el cliente una consultora, la métrica de negocio más adecuada para dirimir la calidad del modelo es ver si efectivamente la inversión se produce en el sector que se predice. EL problema principal al evaluar la calidad del modelo es que esta métrica de negocio muchas veces no se conoce, pues muchas inversiones no son asignadas a un sector concreto y caen bajo el paraguas “*Unallocated*” (desconocido – sin asignar). Otro problema con la calidad de las predicciones es que cabe la posibilidad de que el modelo se vea afectado por el fenómeno conocido como *Data Drift*, dado que con el paso del tiempo se producirá una variación natural en los datos (las inversiones en muchos rubros son cíclicas, por ejemplo, en turismo), por lo que los datos con los que se entrenó el modelo no serán parecidos a los que se les suministra a la hora de ejecutarlo en la vida real. Además, el proceso de colección de datos puede verse afectado por otros problemas, como retardos en la publicación de las notificaciones de las inversiones a los organismos competentes.

Por otra parte, también puede haber un problema en el futuro con el fenómeno conocido como *concept drift*, pues muy posiblemente la clasificación de la variable a predecir (Industry) no permanezca inmutable con el paso del tiempo. Por ejemplo, si bien sectores como el sector primario el del agua y energía parecen a primera vista bastante fáciles de delimitar, la línea que separa una manufactura de un servicio es cada vez más difusa: como señala el *Study on relation industry services productivity*¹⁷. Además, como señalan multitud de estudios, las economías desarrolladas tienden cada vez más a la terciarización en detrimento de sectores como las manufacturas o el primario. Respecto a cómo se producirá este *Concept Drift*, lo normal es que, este cambio se produzca de manera gradual, si bien la llegada de la pandemia del Covid 19 puede acelerar este fenómeno.

¹⁷ Stehrer, R., Baker, P., Foster-McGregor, N., Koenen, J., Leitner, S. M., Schricker, J., ... & Yagafarova, A. (2014). Study on the relation between industry and services in terms of productivity and value creation.



Monitoreo

Para el monitoreo del modelo se utilizará un método basado en ventanas, con una ventana de tiempo lo suficientemente amplia para poder analizar tendencias en las inversiones (seis a doce meses), con un dataset de referencia y el de ventana. Hay que tener en cuenta que el modelo podría verse afectado por el problema conocido como *training service skew*, pues la información de la OCDE incorpora datos de encuestas en sus registros y muchas veces hay valores faltantes (no se declara el valor de la inversión, por ejemplo).

Metodología

Para Sifdi, la metodología más adecuada para llevar a cabo la puesta en producción de su modelo de aprendizaje automático es la metodología tradicional, dado que el tiempo y el presupuesto que se dedican a cada proyecto son variables y los requisitos suelen ser fijos (cada cliente suele requerir el mismo tipo de informe con una periodicidad más o menos fija). La gestión del proyecto se hará de forma interna, lo que permite una buena flexibilidad en la utilización de recursos tanto humanos como económicos. Sifdi se rige por procesos altamente estandarizados (cada informe suele seguir la misma estructura y se suele reciclar mucho contenido de un informe a otro). En este sentido, el proyecto se ejecutará en un ciclo secuencial: inicio, planificación, ejecución, seguimiento/control y cierre, de acuerdo con la metodología PMBOK¹⁸. Así, cada etapa debe completarse para moverse a la siguiente.

¹⁸ <https://www.pmi.org/pmbok-guide-standards>



Conclusiones

El objetivo principal del objetivo era predecir en qué sector se encuadraría una inversión nueva en una de las cinco grandes economías europeas. Nuestro modelo logró superar el *baseline* (el porcentaje de aciertos fue superior al porcentaje de la clase mayoritaria, Servicios dentro del *label*, Industry), pero los resultados a la hora de predecir los sectores en los que se encuadraría una inversión distinta al del rubro de servicios fueron pobres. Respecto al objetivo secundario del trabajo, establecer qué determinantes son más relevantes a la hora de encuadrar una inversión en un sector u otro podemos concluir que dada la debilidad del modelo, no se puede establecer con claridad qué determinantes son más importantes.

Una limitación importante del trabajo es el número de determinantes elegidos para hacer el análisis. Inicialmente se planteó el trabajo con 50 variables repartidas entre los macro-vectores iniciales, pero debido a la gran cantidad de valores faltantes a los que era muy difícil imputar un valor, se optó por incluir solo aquellas variables a las que se podía imputar un valor de manera coherente. Otra limitación importante es el número de operaciones analizadas, pues hay que tener en cuenta que la OCDE recopila información a partir de los datos suministrados por sus miembros, y muchas operaciones no son contabilizadas por los organismos oficiales encargados de analizar los flujos de inversión

Una posible línea de investigación sería usar la información ofrecida por las fuentes *proptech* mencionadas anteriormente para intentar mejorar la capacidad predictiva del modelo. Sin embargo, las Agencias de Promoción de Inversiones son relativamente lentas a la hora de incorporar de incorporar tecnología a sus modelos de negocio. Como señala Jacopo Dettoni¹⁹:

“Históricamente, el sector de promoción de inversiones ha tardado en adoptar la innovación. Las agencias de promoción de inversiones (IPA) de todo el mundo tardaron tiempo en



establecer una presencia sólida en línea (algunas aún se están poniendo al día) y les llevará tiempo aplicar la ciencia de datos a su pensamiento estratégico”.

El *Data Driven Decision Making* ha contribuido a mejorar el rendimiento de infinidad de empresas²⁰: según apuntan Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). utilizando datos de encuestas detallados sobre las prácticas comerciales y las inversiones en tecnología de la información de 179 grandes empresas que cotizan en bolsa, las empresas que adoptan DDD tienen un rendimiento y una productividad entre un 5% y un 6% más altos de lo que se esperaría dadas sus otras inversiones y el uso de tecnología de la información. De hecho, los modelos predictivos aplicados a la localización de inversiones ya han alcanzado un nivel de precisión sin precedentes: tan pronto como Amazon anunció su búsqueda de una segunda sede en América del Norte en septiembre de 2017, dando detalles sobre las características de las comunidades y los activos que le interesaban, los observadores ejecutaron sus modelos y anticiparon la mayoría de las ubicaciones preseleccionadas²¹.

No obstante, muchas decisiones, como la decisión de en qué sector invertir o dónde abrir una nueva filial, pueden regirse en muchas ocasiones por la intuición del HiPPO (*Highest Paid Person's Opinion*) y no tanto en base a factores como los determinantes de la inversión, como en el caso del análisis que nos ocupa en este trabajo. Aunque esta revolución impulsada por los datos está generando un interés sustancial, genera preocupaciones sobre cuestiones como la calidad de los datos, que pueden sesgar las decisiones de inversión para bien o para mal, y plantea dudas sobre el papel que desempeñarán las habilidades personales o 'blandas' en el futuro. de promoción de inversiones y asignación de activos.

²⁰ Brynjolfsson, E., Hitt, L. M., & Kim, H. H. (2011). Strength in numbers: How does data-driven decisionmaking affect firm performance? Available at SSRN 1819486.

²¹ Extraído de <https://www.fdiintelligence.com/article/70913?saveConsentPreferences=success>



Bibliografía

FUENTES ACADÉMICAS

Consultancy UK, December 2019, extraído de <https://www.consultancy.uk/news/23120/three-new-business-models-in-the-consulting-industry>

LaValle, S., Lesser, E., Shockley, R., Hopkins, M. S., & Kruschwitz, N. (2011). Big data, analytics and the path from insights to value. *MIT Sloan management review*, 52(2), 21-32

Jacopo Dettoni, Investment promotion 4.0, Febrero 2018, FDI Intelligence, extraído de <https://www.fdiintelligence.com/article/70913?saveConsentPreferences=success>

McGilvray, D. (2008). *Executing data quality projects: Ten steps to quality data and trusted information (TM)*. Elsevier.

Steinmann, M., Mattei, S. A., & Collmann, J. (2016). A theoretical framework for ethical reflection in big data research. In *Ethical Reasoning in Big Data* (pp. 11-27). Springer, Cham.

Deb Gallagher, Abril 2012, MIT Sloan Review. Extraído de <https://sloanreview.mit.edu/article/the-decline-of-the-hppo-highest-paid-persons-opinion>

Boyd, D., & Crawford, K. (2012). Critical Questions for Big Data in Information, Communication & Society. *Communication and Society*, 15(5), 662-679.

OECD. (1996). *OECD benchmark definition of foreign direct investment*. OECD.

Tenorio, M. A. L., & Sánchez, M. M. L. (2017). Impacto de la Inversión Extranjera Directa en el Desarrollo Económico-Social de la Región. *Jóvenes En La Ciencia*, 2(1), 697-701.

UNCTAD (2020), *World Investment Report 2020: International Production Beyond the Pandemic*, UN, New York, recuperado de <https://doi.org/10.18356/920f7642-en>.



Universidad de Buenos Aires
Facultad de Ciencias Económicas
Escuela de Estudios de Posgrado



EY (junio 2020). Massive decline in foreign direct investment expected in Switzerland and Europe, EY. recuperado de https://www.ey.com/en_ch/news/2020/06/massive-decline-in-foreign-direct-investment-expected-in-switzerland-and-europe

Bruneckiene, J., Jucevicius, R., Zykiene, I., Rapsikevicius, J., & Lukauskas, M. (2019). Assessment of Investment Attractiveness in European Countries by Artificial Neural Networks: What Competences are Needed to Make a Decision on Collective Well-Being?. *Sustainability*, 11(24), 6892.

Bagchi, P., Lejeune, M. A., & Alam, A. (2014). How supply competency affects FDI decisions: some insights. *International Journal of Production Economics*, 147, 239-251

Dornean, A., & Oanea, D. C. (2015). Impact of the economic crisis on FDI in Central and Eastern Europe. *Review of Economic and Business Studies*, 8(2), 53-68.

PÁGINAS WEB CONSULTADAS

<http://sifdi.com/nosotros.html>

<https://cfispain.com/en/quienes-somos/>

<https://iberglobal.com/index.php/quienes-somos>

https://administracion.gob.es/pag_Home/espanaAdmon/Transparencia-y-datos-abiertos

<https://www.boe.es/buscar/doc.php?id=BOE-A-2018-16673>