

ANÁLISIS DEL COMPORTAMIENTO INNOVADOR EN EMPRESAS PERUANAS DE MANUFACTURA

Juan Jesús Arenas

Pontificia Universidad Católica del Perú, Departamento de Ingeniería, Perú
jjarenas@pucp.edu.pe

Miguel Domingo Gonzalez

Pontificia Universidad Católica del Perú, Departamento de Ingeniería, Perú
dgonzal@pucp.edu.pe

Yúser Arturo Cueva

Pontificia Universidad Católica del Perú, Departamento de Ingeniería, Perú
a20131271@pucp.pe

Kervy Dante Rivas

Pontificia Universidad Católica del Perú, Departamento de Ingeniería, Perú
k.rivas@pucp.pe

Paul Aquino Vega

Pontificia Universidad Católica del Perú, Departamento de Ingeniería, Perú
a20121669@pucp.pe

RESUMEN

En 2015 se realizó la Encuesta Nacional de Innovación en la Industria Manufacturera del Perú, la cual incluye preguntas sobre actividades de innovación, recursos humanos, financiamiento para innovación y vinculación con otras entidades. Estos datos contienen información muy valiosa que, mediante técnicas de minería de datos, se podrían describir y servir para futuros trabajos. El objetivo de este artículo es identificar y describir el comportamiento de las empresas para actividades de innovación respecto a los obstáculos presentados en la Encuesta Nacional de Innovación en la Industria Manufacturera del Perú, realizada por el Instituto Nacional de Estadística e Informática, mediante minería de datos. Para ello se utilizará tanto la metodología Cross Industry Standard Process for Data Mining, o también llamada CRISP-DM, así como también las técnicas de coeficiente de Correlación de Pearson y Análisis de Componente Principal, las cuales son usadas en proyectos de minería de datos.

Palabras clave: *minería de datos, innovación empresarial, encuesta de innovación, coeficiente de Correlación de Pearson, Análisis de Componente Principal.*

1. INTRODUCCIÓN

La innovación es un tema crucial para las empresas porque les ayuda a lograr una ventaja competitiva en su sector, así como también a mejorar la economía de la región (Abramo, D'Angelo & Di Costa, 2011; Burnside & Witkin, 2008). Si bien el desarrollo de actividades que generen innovación es responsabilidad de la misma empresa, en la actualidad existen muchos vínculos de cooperación en los cuales se absorbe conocimiento científico y nuevos avances tecnológicos. Dichos vínculos de cooperación se realizan por diferentes objetivos, por ejemplo: asistencia técnica, solicitud de financiamiento, entre otros; y con diferentes agentes productores

de conocimiento y tecnología, tales como universidades, centros de investigación, programas gubernamentales de promoción, etcétera (Branscomb, Kodama & Florida, 1999; Etzkowitz, 2000; Tether, 2002). Cabe resaltar que utilizar la cooperación con instituciones productoras de conocimiento y tecnología es necesario para usar recursos complementarios y obtener nuevos conocimientos, con la finalidad de complementar las actividades de innovación de la empresa (Cassiman & Veugelers, 2002).

Si bien el ámbito científico desarrolla conocimiento y tecnología, no se puede pensar que lo creado en laboratorios será útil para comercializar (Heslop, McGregor & Griffith, 2001). Por ello es necesario realizar otras actividades que aporten a la innovación de la empresa, tales como la adopción de tecnología, cambios incrementales, imitaciones, combinación de conocimientos (Ranga & Etzkowitz, 2013). Estas se obtienen por medio de un vínculo de cooperación (Cohen & Levinthal, 1989; D'Este & Patel, 2007), muchas veces de manera formal: investigaciones conjuntas (Hall, Link & Scott, 2001; Link & Scott, 2005), consultorías (Perkmann & Walsh, 2008) o contratos para proyectos de I+D (Ham & Mowery, 1998); o informal: como interacciones informales (Faulkner & Senker, 1994), movilidad de personal (Zucker, Darby & Torero, 2002) o publicaciones científicas conjuntas (Calvert & Patel, 2003).

Si bien las empresas y los centros tecnológicos cumplen su rol para realizar actividades de innovación, el Estado también tiene funciones de apoyo, fomento e incentivo a los dos agentes mencionados anteriormente. En la actualidad el Estado peruano cuenta con la Ley 30309 de incentivos tributarios (Normas legales, 13 de marzo de 2015) y brinda apoyos económicos para proyectos de desarrollo tecnológico e innovación (Ministerio de la Producción, 2015). También cumple el rol de medir las iniciativas y observar si las actividades de innovación de las empresas están evolucionando y, para ello, mediante el Instituto Nacional de Estadística e Informática del Perú (INEI, 2015), ha realizado la Encuesta Nacional de Innovación en la Industria Manufacturera del Perú para el periodo 2012-2014.

Si bien las encuestas ayudan a obtener resultados estadísticos, existen algunos países que aprovechan sus datos para realizar análisis cuantitativo. Existen investigaciones en España (Segarra-Blasco & Arauzo-Carod, 2008), Italia (Evangelista, Perani, Rapiti & Archibugi, 1997), Reino Unido (D'Este & Patel, 2007), Alemania (Becker & Dietz, 2004) y Brasil (Frank, Cortimiglia, Ribeiro & Oliveira, 2016), donde se han publicado artículos científicos sobre el análisis de las encuestas. Caso contrario sucede en Latinoamérica, donde también existen datos de encuestas, pero no investigaciones científicas con análisis cuantitativo.

Así pues, este artículo tiene como objetivo identificar y describir el comportamiento de las empresas para actividades de innovación respecto a los obstáculos presentados en la Encuesta Nacional de Innovación en la Industria Manufacturera del Perú, utilizando técnicas de minería de datos en la herramienta RStudio. Para ello, este artículo se presenta de la siguiente forma: una introducción a la situación de la innovación en las empresas del Perú; descripción del modelo CRISP-DM, utilizado para minería de datos; desarrollo de dicha metodología, con énfasis en la base de datos utilizada y la forma de presentación de los resultados; descripción de la ejecución del estudio presentando y las variables utilizadas para el análisis, de los resultados de la ejecución del programa en RStudio y su interpretación.

2. METODOLOGÍA

La minería de datos es un conjunto de técnicas de extracción, que organiza, agrupa, relaciona y

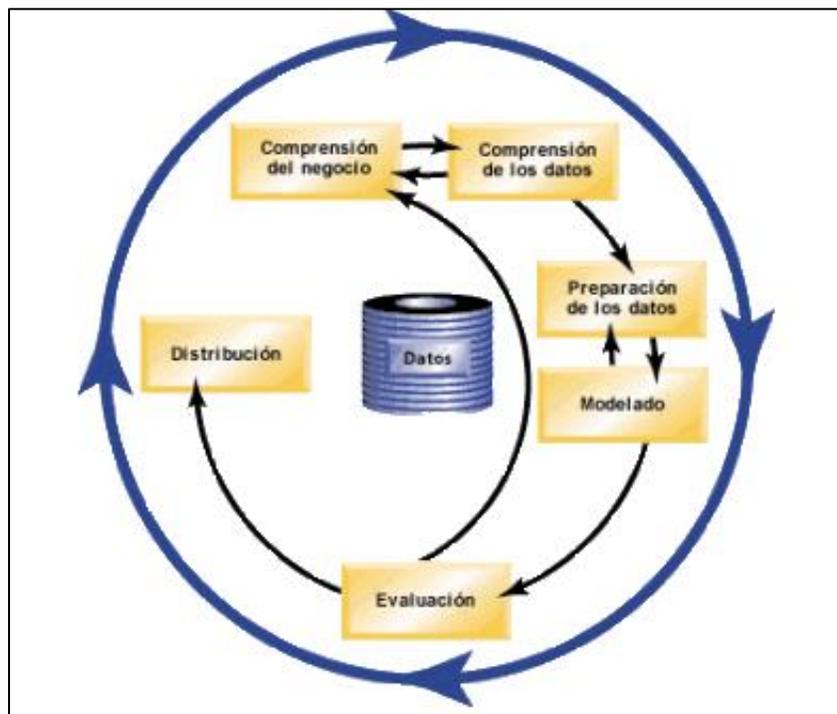
clasifica información de una base de datos (Bozkır, Mazman & Sezer, 2010; Martínez, 2001). Por ejemplo, se utiliza para analizar datos de organizaciones y empresas y encontrar comportamientos sociales (Díaz, Osorio, Amadeo & Romero, 2013).

En este marco se presenta la investigación, que procura conocer el comportamiento de las empresas respecto a sus obstáculos y vínculos para generar innovación, así como también para descubrir otras variables que determinen una información valiosa para el país. La relevancia de este artículo radica en que, mediante las técnicas de minería de datos, es posible definir un comportamiento similar entre empresas que tuvieron los mismos obstáculos y vínculos y así conocer cómo estas se agrupan e identificar qué otras variables son importantes para contar o no con algún vínculo y obstáculo.

2.1. Ejecución del Modelo CRISP-DM para la investigación

CRISP-DM, por sus siglas en inglés *Cross-Industry Standard Process for Data Mining*, es un método utilizado para proyectos de minería de datos, el cual incluye fases, tareas y explicación de las tareas en cada fase de un proyecto de minería de datos. Asimismo, CRISP-DM ofrece un resumen del ciclo de vida del modelo, donde describe la secuencia de fases, aunque no necesariamente la secuencia es escrita debido a que la mayoría de proyectos avanzan y retroceden entre las fases (IBM, 2012).

Figura 1. Ciclo de vida CRISP-DM



Fuente: *Manual CRISP-DM de IBM SPSS Modeler* por IBM, 2012, p. 1.

Entonces, para esta investigación se utilizará este modelo; como primera fase se explicará el negocio. La *comprensión de negocio* tiene la finalidad de definir el objetivo del proyecto y conocer los recursos para el proyecto. En suma, el objetivo principal del proyecto es:

- Analizar el comportamiento innovador de empresas peruanas de manufactura mediante

técnicas de minería de datos.

Como objetivos específicos se presentan los siguientes:

- Identificar los obstáculos más importantes descritos en la encuesta.
- Conocer si existe alguna relación entre los obstáculos que impiden el desarrollo de actividades de innovación en las empresas y los vínculos con instituciones como universidades o centros de investigación.
- Identificar el comportamiento de grupos de empresas por rubros y ciudades sobre la base de los obstáculos que impiden el desarrollo de actividades de innovación en las empresas, vínculos con instituciones, como universidades o centros de investigación, y otras características importantes.

En cuanto a los recursos, se tiene planeado utilizar la herramienta RStudio, la cual está basada en el lenguaje de programación R, muy conocido en minería de datos. Dentro de los recursos también se consideran los datos por utilizar. Para ello se ha usado la Encuesta Nacional de Innovación en la Industria Manufacturera del Perú, realizada en 2015, que se ha complementado con datos generales de la empresa que no se pudieron encontrar en la encuesta.

La siguiente fase del modelo es la **compresión de los datos**, que implica acceder y explorar las bases de datos con el objetivo de conocer y entender los diferentes tipos de datos. Así, para el proyecto se utilizan dos bases de datos:

- **Encuestas de innovación:** son los registros de las 1447 empresas entrevistadas. De ellas, 1068 fueron realizadas a empresas grandes; 48, a medianas; y 331, a pequeñas y microempresas. La encuesta se divide en 11 capítulos: localización de la empresa, identificación de la empresa, actividades de innovación, financiamiento para innovación, encadenamientos e innovación, recursos humanos, resultados de innovación, protección y derechos de propiedad intelectual, fuentes de información y vinculación, obstáculos, e información básica de la empresa.
- **Información general de empresa:** si bien la base de datos anterior contaba con información de la empresa, no fue posible ubicar datos sobre el tipo de empresa ni rubro. Para ello las encuestas se complementaron con información de empresas de la Superintendencia Nacional de Administración Tributaria (SUNAT).

La descripción de los datos se presenta en la siguiente tabla:

Tabla 1. Descripción de los datos

Capítulo	Descripción
Cap. 3: Actividades de innovación	Los resultados fueron descritos con montos en soles invertidos en los tres últimos años (2012, 2013 y 2014) y en diferentes actividades que generan innovación.
Cap. 4: Financiamiento	Los resultados fueron descritos como una distribución porcentual entre los diferentes orígenes de financiamiento.
Cap. 6: Recursos humanos	Los resultados analizados fueron dos: sobre número promedio contratado por la empresa por área y sobre el número promedio contratado por la empresa por profesión.
Cap. 9: Fuentes de innovación	Los resultados están dados por la respuesta —sí o no— para diferentes agentes de

y vinculación	vinculación. Adicionalmente se utiliza la respuesta sí o no para cooperación activa con cada uno de los vínculos.
Cap.10: Obstáculos	Los resultados están dados por la respuesta al grado de importancia (ninguno, bajo, medio, alto).

Fuente: Elaboración propia en base a INEI, 2015.

La tercera fase, **preparación de datos**, es una de las más importantes y tiene como objetivo preparar los datos listos para el modelado. En cuanto a la selección de datos, esto implica que se deben seleccionar las filas (individuos) y las columnas (variables) y crear una matriz (Tan, Steinbach & Kumar, 2013). En el caso del proyecto se crearon dos matrices con diferentes individuos, pero iguales variables. La primera cuenta con los rubros como individuos, por lo que se agrupó a las empresas por rubros de negocio. La segunda cuenta con ciudades como individuos, por lo que también se agrupó a las empresas por ciudades. La descripción de las variables se muestra en la tabla 2:

Tabla 2. Descripción de los datos

Capítulo	Variable	Tipo
Actividades de innovación (Capítulo 3)	C3P1_1 - Actividades I+D interna C3P1_2 - Actividades I+D Externa C3P1_3 - Adquisición de bienes de capital C3P1_4 - Adquisición de <i>hardware</i> C3P1_5 - Adquisición de <i>software</i> C3P1_6 - Transferencia de tecnología C3P1_7 - Diseño e ingeniería industrial C3P1_8 - Capacitación para actividades de innovación C3P1_9 - Estudios de mercado para introducción de innovaciones	Suma de montos por año por actividad. Valor: número decimal positivo.
Financiamiento (Capítulo 4)	C4P1_1 - Apoyos gubernamentales C4P1_2 - Banca comercial privada C4P1_3 - Otras empresas C4P1_4 - Recursos propios C4P1_5 - Otras fuentes	Suma del porcentaje (0-100). Número entero positivo.
Recursos Humanos (Capítulo 6)	C6P2_1 - Ingenierías y tecnologías C6P2_2 - Ciencias naturales y exactas C6P2_3 - Ciencias médicas y de la salud C6P2_4 - Ciencias agrarias C6P2_5 - Ciencias sociales C6P2_6 - Humanidades C6P3_1 - Área de Informática y Sistemas C6P3_2 - Área de Investigación y Desarrollo C6P3_3 - Área de Ingeniería Industrial C6P3_4 - Área de Gestión de la Calidad	Suma del promedio (0-100). Número entero positivo.
Fuentes de información y vinculación (Capítulo 9)	C9P2_1 - Universidades C9P2_2 - Institutos de investigación públicos C9P2_3 - Institutos de investigación privados C9P2_4 - Programas gubernamentales de promoción de CTI C9P2_5 - Laboratorios no universitarios C9P2_6 - CITE C9P2_7 - SENATI y otras de vinculación técnica C9P2_8 - Proveedores C9P2_9 - Clientes C9P2_10 - Casa matriz u otras empresas del grupo C9P2_11 - Competidores u otras empresas C9P2_12 - Gremios empresariales	El resultado fue dado como respuesta sí o no, por lo que se sumaron los valores con respuesta positiva (sí). Número entero positivo.

	C9P2_13 - Consultores y expertos	
Obstáculos (Capítulo 10)	C10P1_1 - Escasez de personal calificado C10P1_2 - Insuficiente información sobre tecnologías C10P1_3 - Insuficiente información sobre los mercados C10P1_4 - Dificultades para encontrar socios de cooperación para la innovación C10P1_5 - Mercado dominado por empresas establecidas C10P1_6- Incertidumbre respecto a la demanda de bienes y servicios innovadores C10P1_7 - Reducido tamaño del mercado C10P1_8 - La innovación es fácil de imitar C10P1_9 - Falta de fondos en la empresa o grupo de empresas C10P1_10 - Falta de financiación de fuentes exteriores a la empresa C10P1_11 - La innovación tiene un costo demasiado elevado C10P1_12 - Percepción de riesgos económicos excesivos C10P1_13 - Rigidez organizativa dentro de la empresa C10P1_14 - Insuficiente flexibilidad de los reglamentos o normas C10P1_15 - Infraestructura física inadecuada C10P1_16 - Limitaciones de las políticas públicas de ciencia y tecnología y su instrumento C10P1_17 - Otros	El resultado fue dado como respuesta de ninguno, bajo, medio o alto. Se los cuantificó como ninguno=3, bajo=-1, medio=-2, alto=-3, por lo que se sumaron los valores. Número entero negativo o positivo.

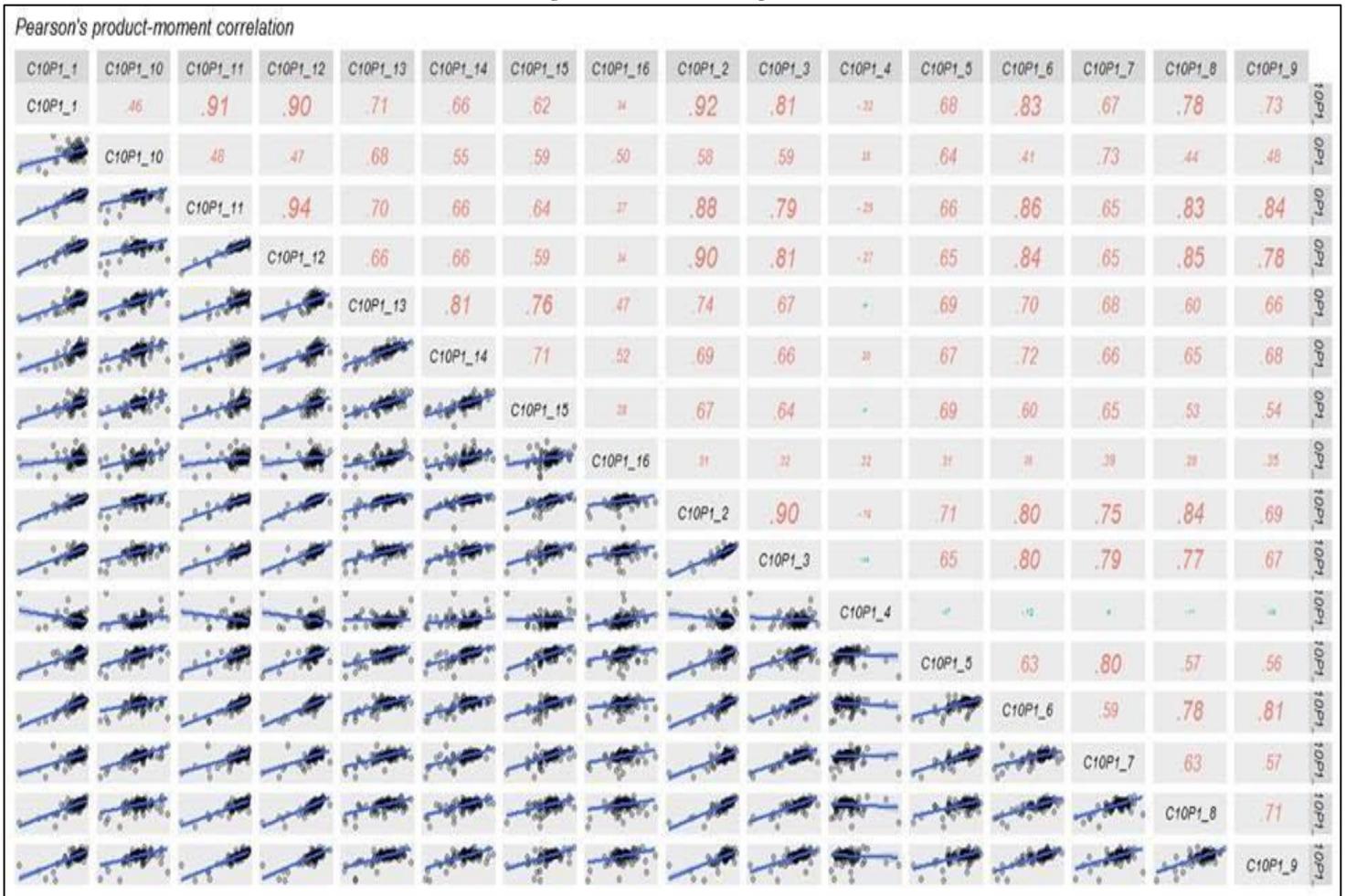
Fuente: Elaboración propia en base a INEI, 2015.

La cuarta fase, **modelado**, es el conjunto de actividades que sirven para definir las técnicas necesarias que ayuden a cumplir el objetivo. Normalmente, los analistas de datos ejecutan varios modelos utilizando parámetros por defecto y los ajustan o vuelven a la fase de preparación para las manipulaciones necesarias. Para el proyecto, luego se muchas iteraciones, se determinó realizar el análisis con los siguientes modelos:

- **Coefficiente de Correlación de Pearson (CCP)** (Strychalska-Rudzewicz, 2015): es una técnica que ayuda a determinar la relación entre dos variables. El grado de coeficiente es un valor entre -1 a 1, siendo más fuerte la relación directa cuando el valor se acerca a 1 y más fuerte inversa cuando el valor se acerca a -1. El coeficiente de correlación también se puede mostrar por medio del ángulo formado entre dos variables. Así pues, el ángulo más cerca de 0 es como una relación directa y un ángulo cerca de 180 grados, una relación inversa. En el caso de que el ángulo entre dos variables sea cerca de 90, no se podrá encontrar relación.
- **Análisis de componentes principales (ACP)**: se utiliza con la finalidad de reducir el número de variables independientes en un modelo explicativo (Jolliffe, 2002). El método ACP establece un procedimiento para reducir un set de «componentes» que describen una porción conocida de la variación total en las variables originales (*principales of principal*).

La quinta fase es la **evaluación**. En esta etapa deben evaluarse los modelos presentados en la fase anterior. Las conclusiones o interferencias obtenidas de los modelos y procesos de minería de datos reciben el nombre de *descubrimiento*. Debido a la gran cantidad de variables, en este capítulo se utiliza la técnica de CCP con el objetivo de encontrar algunas variables que representen a las demás y puedan ser comparadas con los otros capítulos. En la figura 2 se puede apreciar que la variable C10P1_8 «La innovación es fácil de imitar» cuenta con un coeficiente mayor a 0.70 con casi todas las variables, lo que significa que las empresas que expresaron contar con ese obstáculo también dijeron mostrar otros, por lo que se ha determinado que C10P1_8 será la variable que represente al capítulo 10.

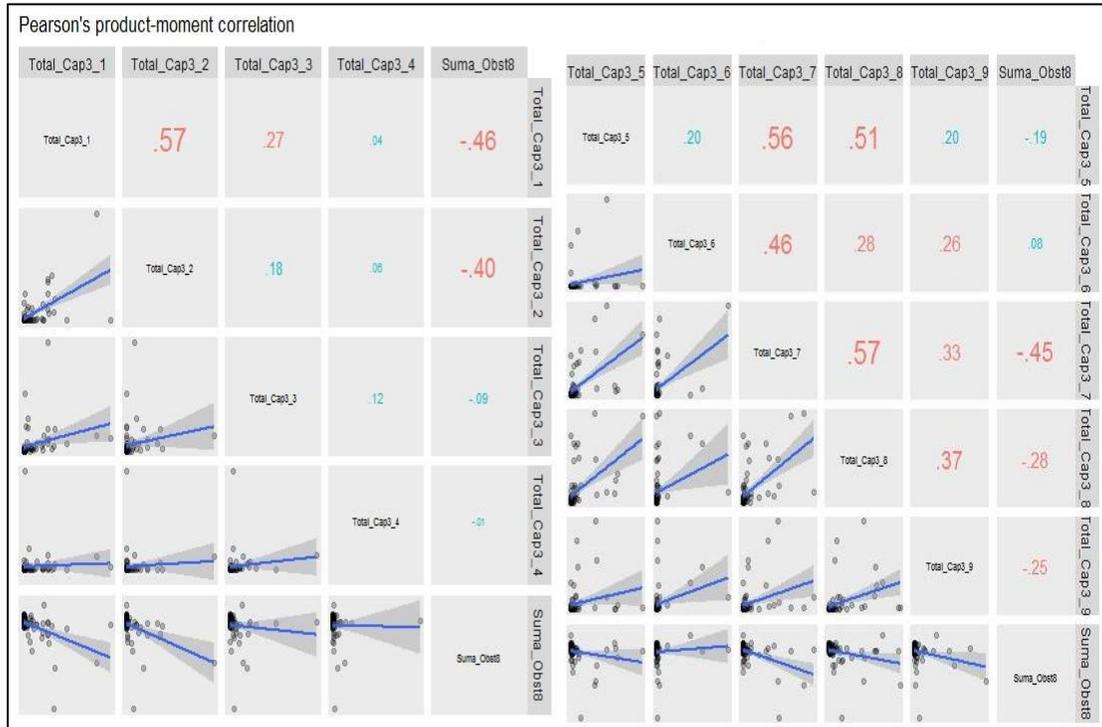
Figura 2. CPP del Cap10



Fuente: Elaboración propia en base a INEI, 2015.

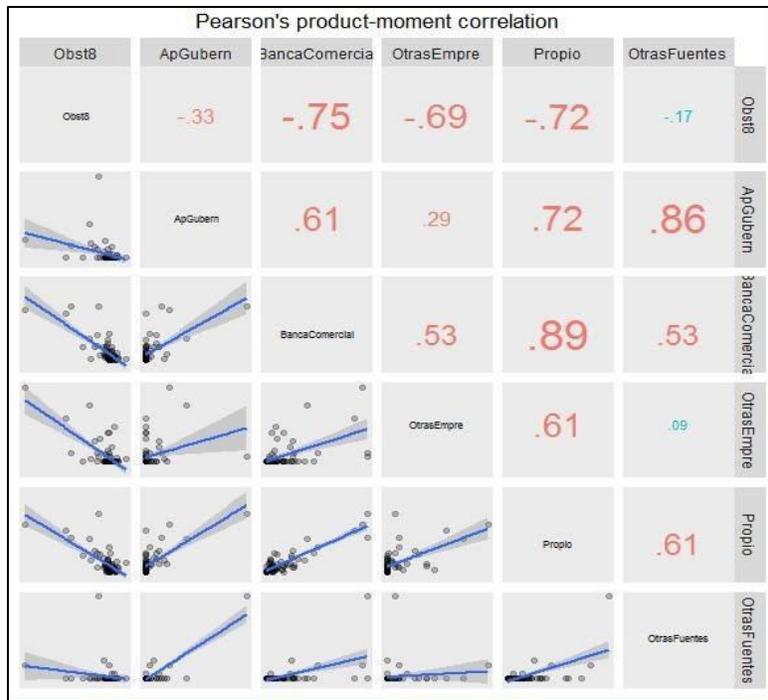
Luego de definir al representante del capítulo 10 (C10P1_8), se utilizará la técnica CCP entre las demás variables de los capítulos con C10P1_8. La figura 3 muestra el resultado de la aplicación de la técnica para la variable C10P1_8 y las variables del capítulo 3. La figura 4 muestra la técnica CCP para C10P1_8 y capítulo 4. La figura 5 muestra la técnica CCP para C10P1_8 y capítulo 6. Por último, La figura 6 muestra la técnica CCP para C10P1_8 y capítulo 9.

Figura 3. CPP del Cap3 y C10P1_8



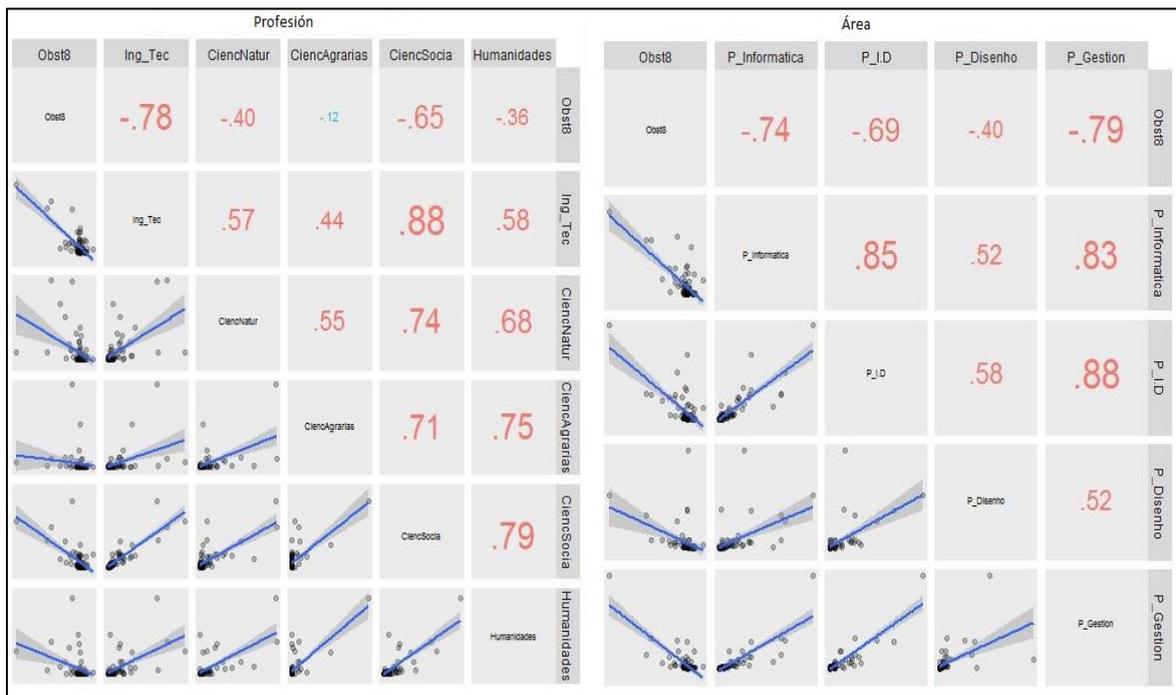
Fuente: Elaboración propia en base a INEI, 2015.

Figura 4. CPP del Cap4 y C10P1_8



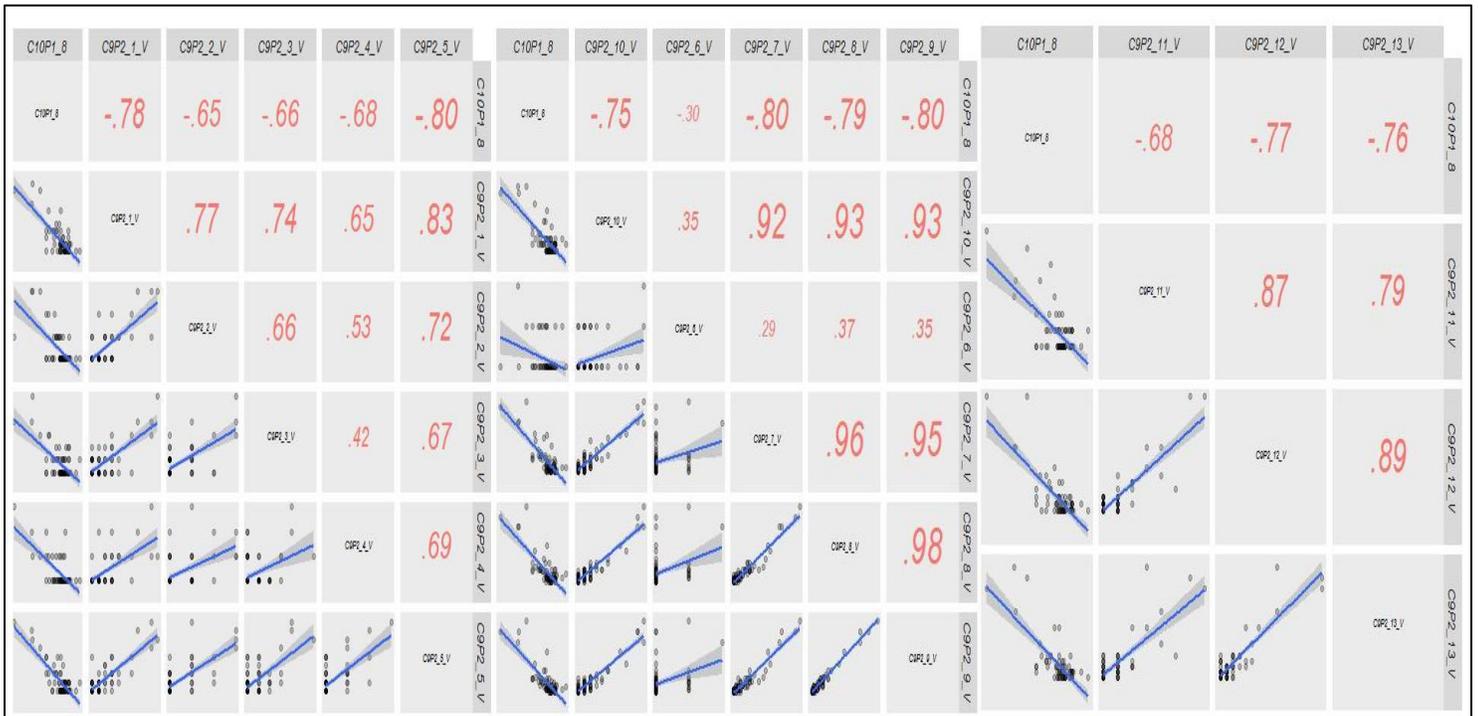
Fuente: Elaboración propia en base a INEI, 2015.

Figura 5. CPP del Cap6 y C10P1_8



Fuente: Elaboración propia en base a INEI, 2015.

Figura 6. CPP del Cap9 y C10P1_8

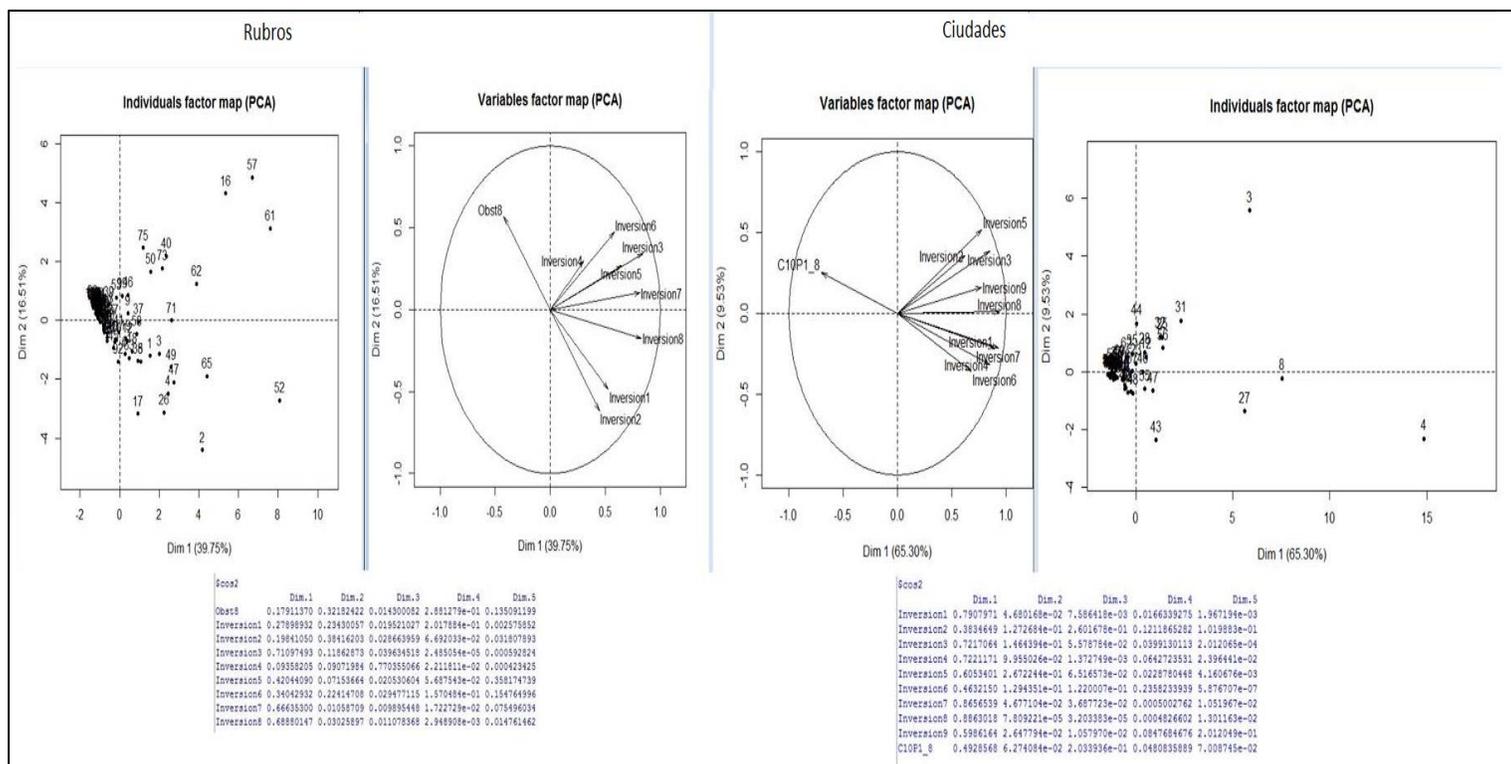


Fuente: Elaboración propia en base a INEI, 2015.

Luego de evaluar la relación que existe entre los obstáculos y los demás capítulos, se mostrarán los resultados obtenidos por la técnica ACP. A diferencia de los gráficos mostrados por RStudio para la técnica CCP, los gráficos de ACP exponen tres imágenes. En la parte superior izquierda se muestra la representación de los individuos; en la parte superior derecha, la representación de las variables. Este gráfico también puede ayudar a conocer la relación entre variables, siendo el ángulo más cercano a 0 una relación directa; si es 90, no existe relación; y si es 180, hay una relación indirecta. En la parte inferior se muestra el grado de representación de las variables, siendo la suma de la dimensión 1 y dimensión 2 un valor entre 0 y 1; de ahí que un valor más cercano a 1 resulte como mejor representado. También se puede apreciar el grado de representación de todo el grafico en la suma de los ejes.

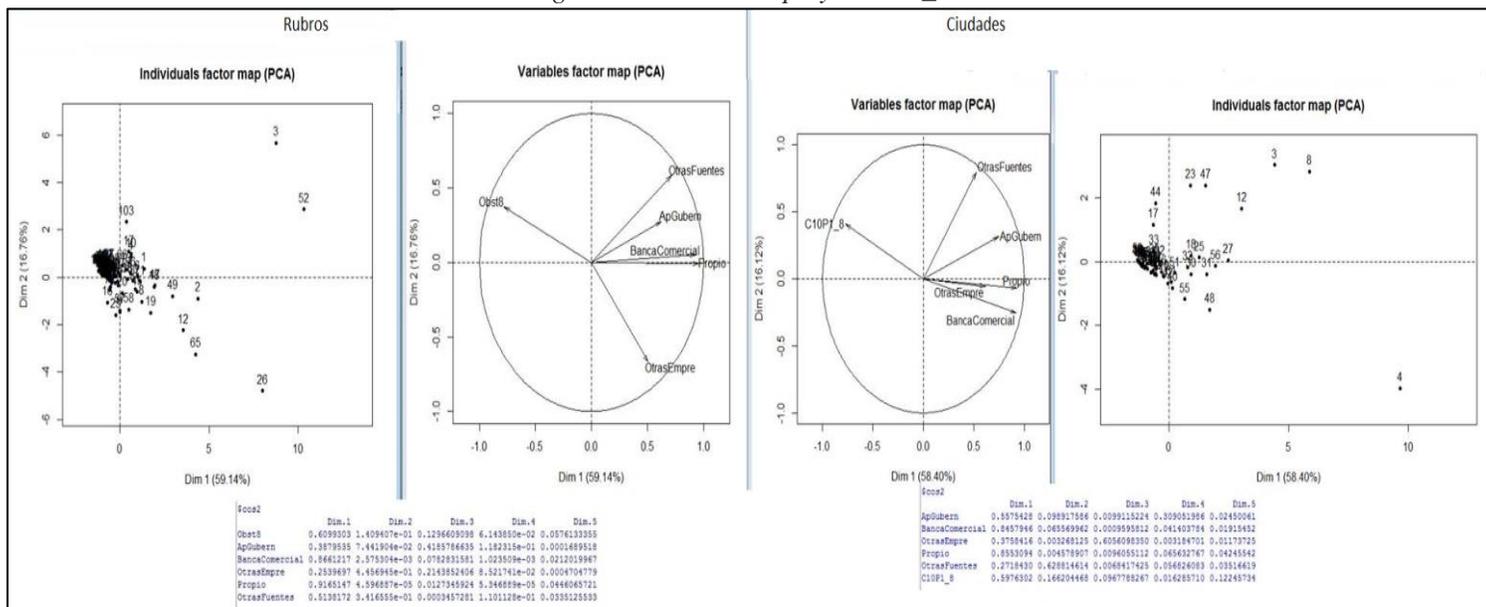
En las siguientes figuras se muestra el análisis para individuos como rubros y ciudades. En la figura 7 se muestra el análisis del capítulo 3 y el obstáculo C10P1_8. En la figura 8 se muestra el análisis del capítulo 4 y el obstáculo C10P1_8. En la figura 9 se muestra el análisis de la variable profesiones correspondiente al capítulo 6 y el obstáculo C10P1_8. En la figura 10 se muestra el análisis de la variable áreas correspondiente al capítulo 6 y el obstáculo C10P1_8. Por último, En la figura 11 se muestra el análisis correspondiente al capítulo 9 y el obstáculo C10P1_8

Figura 7. ACP del Cap3 y C10P1_8



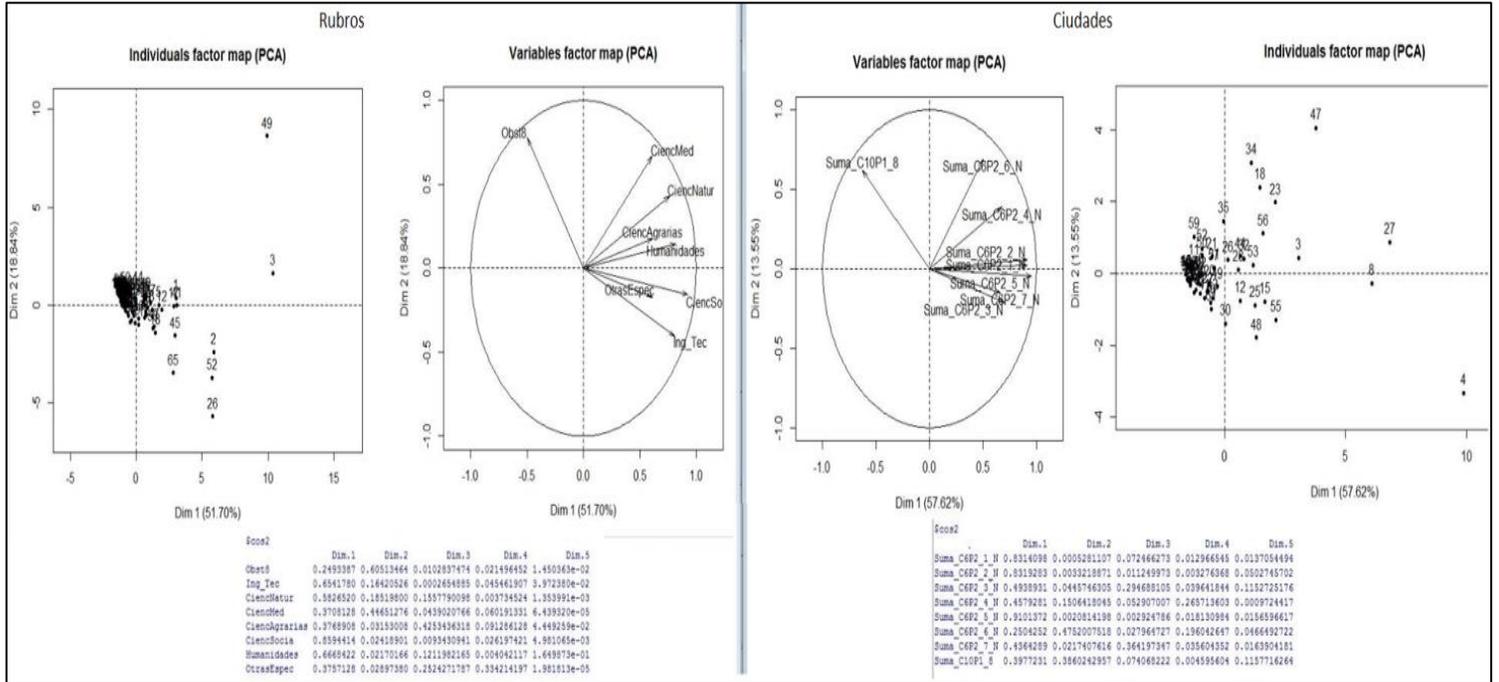
Fuente: Elaboración propia en base a INEI, 2015.

Figura 8. ACP del Cap4 y C10P1_8



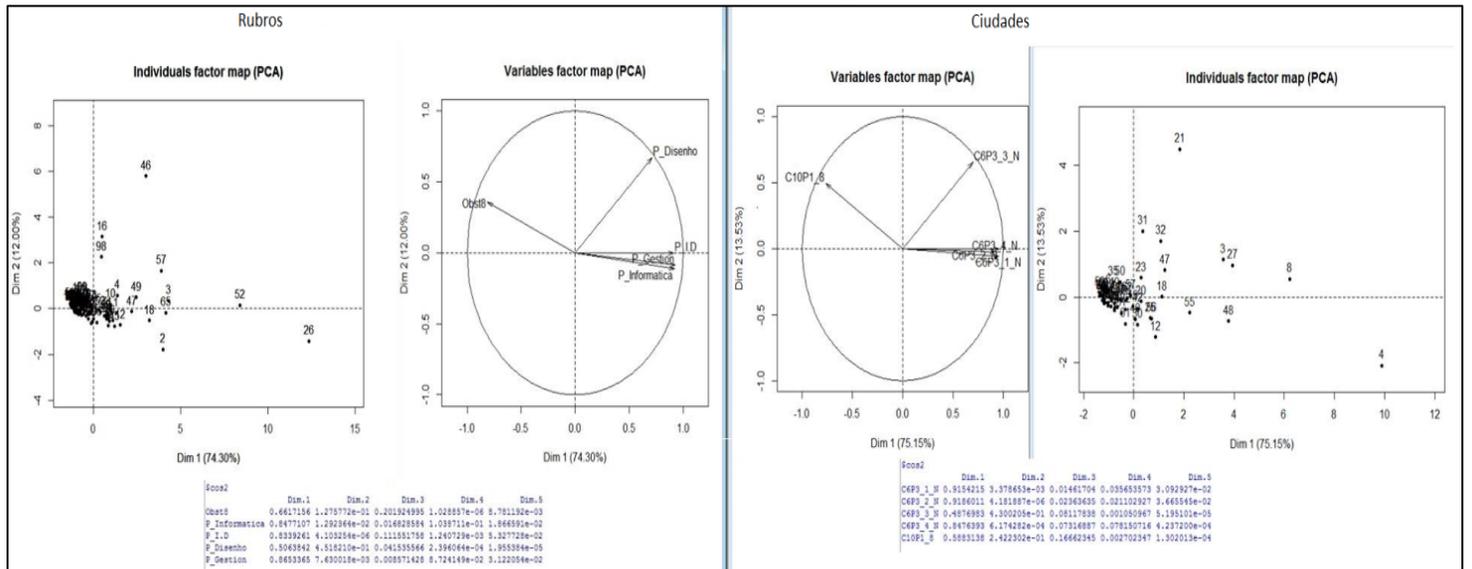
Fuente: Elaboración propia en base a INEI, 2015.

Figura 9. ACP del Cap6-profesión y C10P1_8



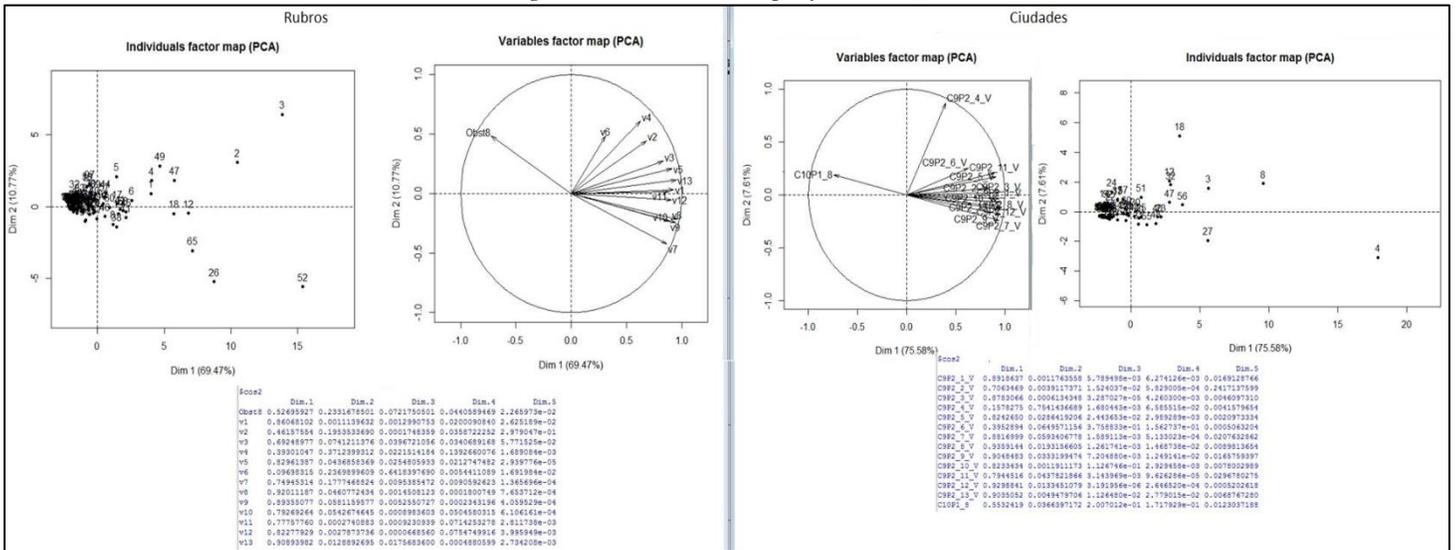
Fuente: Elaboración propia en base a INEI, 2015.

Figura 10. ACP del Cap6-Área y C10P1_8



Fuente: Elaboración propia en base a INEI, 2015.

Figura 11. ACP del Cap9 y C10P1_8



Fuente: Elaboración propia en base a INEI,2015.

3. RESULTADOS

Una vez que se ha terminado con la evaluación, se desarrolla la última fase: **distribución**. Esta sirve para mostrar los resultados referentes a la comprensión del negocio. En esta etapa se puede realizar un resumen de lo encontrado en lo que respecta al entendimiento del negocio. Para este artículo se resumirá el obstáculo representativo y se verá cómo las empresas lo superan.

Al respecto, se puede concluir lo siguiente:

1. El obstáculo C10P1_8 «la innovación fácil de imitar» es uno de los obstáculos más representativos, debido a que se relaciona directamente con casi todos los demás obstáculos. Ello es posible al observar que aquellas empresas que han presentado el obstáculo «la innovación es fácil de imitar» también tuvieron un impacto en obstáculos, como, por ejemplo: C10P1_1 «Escases de personal calificado» o C10P1_11 «la innovación tiene un costo demasiado alto», entre otros.
2. El capítulo 3 «actividades para innovación» no genera un valor significativo para superar o contribuir al grado de impacto del obstáculo representativo C10P1_8.
3. Según el capítulo 4 «tipos de financiamiento para la innovación», aquellas empresas que buscaron financiamiento en C4P1_2 «Banca Comercial Privada», C4P1_3 «Otras empresas» y C4P1_4 «Recursos propios» pudieron superar el obstáculo representativo. Asimismo, aquellas empresas que buscaban financiamiento en C4P1_1 «apoyo gubernamental» hacían lo mismo en C4P1_5 respecto a «otras fuentes» y que aquellas que buscaban en C4P1_2 «Banca Comercial Privada» hacían lo propio en C4P1_4 «Recursos propios» y viceversa.
4. De acuerdo con el capítulo 6 «Recursos humanos», aquellas empresas que contaban con profesionales como C6P2_1 «Ingeniería y tecnología» y C6P2_5 «Ciencias Sociales» o áreas como C6P3_1 «Áreas de informática o sistemas», C6P3_2 «Áreas de I+D» y C6P3_4 «Áreas de gestión de la calidad» superaron el obstáculo representativo.

5. Según el capítulo 9 «vinculación para la innovación», aquellas empresas que se vincularon con C9P2_1 «Universidades», C9P2_5 «Laboratorios no universitarios», C9P2_7 «SENATI y otras vinculaciones técnicas», C9P2_8 «Proveedores», C9P2_9 «Clientes», C9P2_10 «Casa matriz u otro grupo de empresas del grupo», C9P2_12 «Gremios empresariales» y C9P2_13 «Consultores y expertos» pudieron superar el obstáculo representativo. También se aprecia que las empresas que se vincularon con C9P2_7 «SENATI y otras vinculaciones técnicas» también se vincularon con C9P2_8 «Proveedores» y C9P2_9 «Clientes».
6. De acuerdo con la técnica de ACP y el análisis con los capítulos 3, las empresas constituidas en los distritos de Lima (Ate, San Isidro y Santa Anita) y la ciudad de Lambayeque, que han realizado actividades de C3P1_1 «I+D interna», C3P1_4 «Adquisiciones de *hardware*», C3P1_6 «transferencia tecnológica» y C3P1_7 «Diseño e ingeniería industrial», han superado el obstáculo representativo.
7. Según la técnica de ACP y el análisis con los capítulos 4, las empresas constituidas en los distritos de Lima (Ate, Los Olivos, Lurigancho y San Juan de Lurigancho), que han buscado financiamiento en C4P1_2 «Banca Comercial Privada», han superado el obstáculo representativo. También se evidencia que las empresas cuyos rubros son: 12 «Elaboración de otros productos alimenticios», 19 «Tejedura de productos textiles», 26 «Fabricación de prendas de vestir», 65 «Fabricación de productos metálicos para uso estructural», que han buscado financiamiento en C4P1_3 «Otras Empresas», han superado el obstáculo representativo.
8. Según la técnica de ACP y el análisis con los capítulos 6, las empresas constituidas en los distritos de Lima (Ate, Chorrillos, El Agustino, La Victoria, San Juan de Lurigancho y Santa Anita), que cuentan profesionales en «Ciencias Médicas y Salud», han superado el obstáculo representativo. También se evidencia que las empresas cuyos rubros son: 2 «Elaboración y conservación de pescado, crustáceos y moluscos», 26 «Fabricación de prendas de vestir», 45 «Fabricación de otros productos químicos», 52 «Fabricación de productos de plástico» y 65 «Fabricación de productos metálicos para uso estructural», que cuentan profesionales en C6P2_1 «Ingeniería y Tecnología», han superado el obstáculo representativo.
9. De acuerdo con la técnica de ACP y el análisis con los capítulos 6, las empresas constituidas en los distritos de Lima (Ate, Chorrillos, San Juan de Lurigancho y Santa Anita) y la ciudad de Lambayeque, que cuentan con áreas de Informática, Gestión de la Calidad o I+D, han superado el obstáculo representativo. También se evidencia que las empresas cuyos rubros son: 18 «Hilatura, tejedura y acabado de productos textiles», 26 «Fabricación de prendas de vestir», 47 «Fabricación de otros productos químicos», 52 «Fabricación de productos de plástico» y 65 «Fabricación de productos metálicos para uso estructural», que cuentan con áreas en Informática o I+D, han superado el obstáculo representativo.
10. Según la técnica de ACP y el análisis con los capítulos 9, las empresas constituidas en los distritos de Lima (Ate, San Juan de Lurigancho y Santa Anita) y la ciudad de Lambayeque, que cuentan con vinculación con «SENATI y otros centros técnicos» o «Proveedores», han superado el obstáculo representativo. También se evidencia que las empresas cuyos rubros son: 18 «Hilatura, tejedura y acabado de productos textiles», 26 «Fabricación de prendas de vestir», 52 «Fabricación de productos de plástico» y 65 «Fabricación de productos metálicos para uso estructural», que cuentan con «SENATI y otros centros técnicos» o «Proveedores»,

han superado el obstáculo representativo.

4. CONCLUSIONES

En la actualidad, el uso de análisis de datos ha tomado mucha importancia debido a la gran cantidad de datos almacenados en sistemas de información. En nuestro caso, el universo de datos está basado en una encuesta, la cual fue complementada con información general de empresas, obtenida de los sistemas de información de la SUNAT.

En cuanto a su importancia, hay que resaltar que este estudio puede colocar a la minería de datos como un conjunto de técnicas que ayuden a obtener comportamientos que no son fáciles de encontrar en las empresas. Respecto a la innovación, esta es sustancial porque muestra la situación de un país respecto a una muestra de datos relevantes.

Sobre lo obtenido se puede concluir que las empresas tienen muchos problemas para innovar, de las cuales se aprecia una lista de obstáculos que las empresas han colocado como las más importantes. Luego de las técnicas de minería de datos, se observan obstáculos importantes de análisis, los cuales muestran una correlación muy alta entre otros obstáculos. Para este artículo se seleccionó el obstáculo “la innovación es fácil de imitar” aunque existen otros obstáculos (por ejemplo: “insuficiente información de la tecnología” o “insuficiente información sobre el mercado”) que serían interesantes para futuras investigaciones.

El análisis de minería de datos también muestra que contar con profesionales en informática y ciencias sociales o pertenecer a los rubros «fabricación de textiles», «fabricación de productos de plástico» y «fabricación de productos metálicos» puede ayudar a mitigar que los proyectos innovadores fueran fáciles de imitar.

Por último, la minería de datos da luces que existen empresas que vienen trabajando de forma colaborativa y les está permitiendo mitigar el obstáculo en análisis. Se aprecia que distritos de Lima —como Ate, Santa Anita y San Juan de Lurigancho— y la ciudad de Lambayeque están realizando actividades que le permiten crear proyectos innovadores difíciles de imitar.

REFERENCIAS

- Abramo, G., D'Angelo, C. A., & Di Costa, F. (2011). University-industry research collaboration: a model to assess university capability. *Higher Education*, 62(2), 163-181. doi: 10.1007/s10734-010-9372-0
- Becker, W., & Dietz, J. (2004). R&D cooperation and innovation activities of firms - Evidence for the German manufacturing industry. *Research Policy*, 33(2), 209-223. doi: 10.1016/j.respol.2003.07.00
- Bozkır, A. S., Mazman, S. G., & Sezer, E. A. (2010). Identification of user patterns in social networks by data mining techniques: Facebook case. *Technological Convergence and Social Networks in Information Management*, 145-153. doi: 10.1007/978-3-642-16032-5_13
- Branscomb, L. M., Kodama, F., & Florida, R. (1999). *Industrializing Knowledge. University-Industry Linkages in Japan and the United States*. Cambridge: The MIT Press.
- Burnside, B., & Witkin, L. (2008). Forging successful university-industry collaborations. *Research Technology Management*, 51(2), 26-30. Recuperado de <http://search.proquest.com/docview/213802791?accountid=28391>
- Cassiman, B., & Veugelers, R. (2002). R&D Cooperation and Spillovers: Some empirical evidence from Belgium. *The American Economic Review*, 92(4), 1169-1184. doi: 10.1257/00028280260344704

- Calvert, J., & Patel, P. (2003). University-industry research collaborations in the UK: bibliometric trends. *Science and Public Policy*, 30(2), 85–96.
- Cohen, W. M., & Levinthal, D. A. (1989). Innovation and learning: the two faces of R&D. *The Economic Journal*, 99(397), 569-596. doi: 10.2307/2233763
- D’Este, P. D., & Patel, P. (2007). University-Industry linkages in the UK: What are the factors underlying the variety of interactions with industry? *Research Policy*, 36(9), 1295-1313. doi: 10.1016/j.respol.2007.05.002
- Díaz, F. J., Osorio, M. A., Amadeo, A. P., & Romero, D. (2013). Aplicando estrategias y tecnologías de Inteligencia de Negocio en sistemas de gestión académica. En C. Caluva, S. Aranguren & R. Muzachiodi (Comps.), *XV Workshop de Investigadores en Ciencias de la Computación* (pp. 225-229). La Plata: Universidad Autónoma de Entre Ríos. Recuperado de <http://sedici.unlp.edu.ar/handle/10915/27157>
- Etzkowitz, H. (2000). Tech transfer, incubators probed at Triple Helix III. *Research Technology Management*, 43(6), 4-5. Recuperado de <http://search.proquest.com/docview/213806468?accountid=28391>
- Evangelista, R., Perani, G., Rapiti, F., & Archibugi, D. (1997). Nature and impact of innovation in manufacturing industry: some evidence from the Italian innovation survey. *Research Policy*, 26(4-5), 521-536. doi: 10.1016/S0048-7333(97)00028-0
- Faulkner, W., & Senker, J. (1994). Making sense of diversity: public-private sector research linkage in three technologies. *Research Policy*, 23(6), 673–695. [https://doi.org/10.1016/0048-7333\(94\)90017-5](https://doi.org/10.1016/0048-7333(94)90017-5)
- Frank, A. G., Cortimiglia, M. N., Ribeiro, J. L. D., & Oliveira, L. S. de. (2016). The effect of innovation activities on innovation outputs in the Brazilian industry: Market-orientation vs. technology-acquisition strategies. *Research Policy*, 45(3), 577-592. doi: 10.1016/j.respol.2015.11.011
- Hall, B. H., Link, A. N., & Scott, J. T. (2001). Barriers Inhibiting Industry from Partnering with Universities: Evidence from the Advanced Technology Program. *The Journal of Technology Transfer*, 26(1), 87–98. <https://doi.org/10.1023/a:1007888312792>
- Ham, R. M., & Mowery, D. C. (1998). Improving the effectiveness of public–private R&D collaboration: case studies at a US weapons laboratory. *Research Policy*, 26(6), 661–675. [https://doi.org/10.1016/S0048-7333\(97\)00041-3](https://doi.org/10.1016/S0048-7333(97)00041-3)
- Heslop, L. A., McGregor, E., & Griffith, M. (2001). Development of a Technology Readiness Assessment Measure: The Cloverleaf Model of Technology Transfer. *Journal of Technology Transfer*, 26(4), 369-384. doi: 10.1023/A:1011139021356
- IBM. (2012). *Manual CRISP-DM de IBM SPSS Modeler*. Nueva York: Autor. Recuperado de <ftp://public.dhe.ibm.com/software/analytics/spss/documentation/modeler/15.0/es/CRISP-DM.pdf>
- INEI. (2015). Encuesta Nacional de Innovación en la industria manufacturera. Lima, Peru: Dirección Nacional de Censos y Encuestas.
- Jolliffe, I. (2002). *Principal component analysis* (2ª ed.) Nueva York: Springer.
- Martínez, G. (2001). Minería de datos. Cómo hallar una aguja en un pajar. *Ingenierías*, XIV, 53, 53-66.
- Ministerio de la Producción. (2015). *Innovate: Caja de herramientas para la innovación*. Lima: Autor. Recuperado de <http://www.innovateperu.pe/media/innovateperu/libro-digitse.pdf>
- Normas legales. (13 de marzo de 2015). *El Peruano*, pp. 54844-54846. Recuperado de <http://www.leyes.congreso.gob.pe/Documentos/Leyes/30309.pdf>
- Perkmann, M., & Walsh, K. (2008). Engaging the scholar: Three types of academic consulting and their impact on universities and industry. *Research Policy*, 37(10), 1884–1891. <https://doi.org/10.1016/j.respol.2008.07.009>
- Ranga, M., & Etzkowitz, H. (2013). Triple Helix systems: an analytical framework for innovation policy and practice in the Knowledge Society. *Industry and Higher Education*, 27(4), 237-262. doi: 10.5367/ihe.2013.0165
- Segarra-Blasco, A., & Arauzo-Carod, J. M. (2008). Sources of innovation and industry-university interaction:

Evidence from Spanish firms. *Research Policy*, 37(8), 1283-1295. doi: 10.1016/j.respol.2008.05.003

Strychalska-Rudzewicz, A. (2015). Cultural dimensions and innovation. *Socio-Economic Problems and the State*, 13(2), 59-67. Recuperado de <http://web.a.ebscohost.com/ehost/detail/detail?sid=ba1c11c0-72f1-4b1f-9b2b-b662a713a46c%40sessionmgr4008&vid=0&hid=4212&bdata=JnNpdGU9ZWZWhvc3QtbGl2ZQ%3d%3d#AN=111329397&db=a9h>

Tan, P.-N., Steinbach, M., & Kumar, V. (2013). Data. En *Introduction to Data Mining* (65-72). Boston: Addison-Wesley. doi: 10.1016/j.cll.2007.10.008

Tether, B. S. (2002). Who co-operates for innovation, and why: An empirical analysis. *Research Policy*, 31(6), 947-967. doi: 10.1016/S0048-7333(01)00172-X

Zucker, L., Darby, M., & Torero, M. (2002). Labor Mobility from Academe to Commerce. *Journal of Labor Economics*, 20(3), 629–660. <https://doi.org/10.1086/339613>