

UTILIZACIÓN DE HERRAMIENTAS DE INTELIGENCIA DE NEGOCIOS PARA EVALUAR EL COMPORTAMIENTO DE INDICADORES DE MEDICIÓN DE GESTIÓN ORGANIZACIONAL (KPI) PARA SER APLICADOS A FUENTES MASIVAS DE DATOS FINANCIEROS

CAMPANARO, ROSA S.

rcampanaro@fcecon.unr.edu.ar

DEMARTIS, PABLO NICOLAS

ndemartis@fcecon.unr.edu.ar

DIAZ, DANIEL J.

ddiaz@fcecon.unr.edu.ar

DIAZ TOLEDO, SELVA

selvadiaztoledo@hotmail.com

VIOLA, MONICA B.

monicaviola@arnet.com.ar

Facultad de Ciencias Económicas y Estadística. Universidad Nacional de Rosario

ÁREA TEMÁTICA B: TRABAJOS DE INVESTIGACIÓN.

PALABRAS CLAVE: XBRL - Análisis de Informes Financieros - KPI (Indicadores Claves de Performance) - Data Science (Ciencia de Datos) - Big Data (fuentes masivas de datos).

RESUMEN.

Dos claras líneas de desarrollo se imponen actualmente, en el ámbito de la tecnología informática: el uso de “Big Data (fuentes masivas de datos)” y “Data Science (ciencia de datos)”. Esta tendencia, prácticamente irreversible, crea la necesidad de evaluar el impacto que estos desarrollos tendrán sobre las técnicas, modelos y prácticas arraigadas en la gestión de organizaciones.

La investigación en curso, tiene como base el uso de cantidades masivas de datos financieros extraídos de la SEC (Securities and Exchange Commission) en formato XBRL. Sobre esa base, en la etapa previa del proyecto de investigación, se analizaron modelos de predictibilidad de insolvencia y quiebras empresariales.

El objetivo de la presente etapa es poder estudiar, sobre esa misma fuente masiva de datos financieros, la pertinencia de modelos de evaluación de performance de gestión. Para ello nos centraremos en el estudio bibliográfico, de Indicadores Claves de Performance – KPI (Key Performance Indicators). También se prevé analizar la pertinencia de aplicar técnicas de Ciencia de Datos, como soporte a dichos modelos.

El presente trabajo describe objetivos, lineamientos, metodología, resultados esperados y aspectos relevantes del proyecto de investigación, en lo pertinente a la selección de KPIs a aplicar sobre esa fuente de datos financieros de empresas.

1. INTRODUCCIÓN.

El año anterior, se desarrolló un trabajo que expone el avance de proyecto hasta ese momento, y en el que se pudo visualizar consideraciones relativas a la evaluación de modelos de predictibilidad de quiebras empresariales al ser aplicados a cantidades masivas de Datos financieros extraídos de la SEC - Securities and Exchange Commission. (Campanaro R. y otros - 2016)¹.

Del análisis de resultados obtenidos de la primera etapa del proyecto, se observó, que si bien los modelos de predictibilidad de quiebras sirven para pronosticar con un grado razonable de certeza la posibilidad de estados de insolvencia empresarial, no nos brindan una herramienta óptima para evaluar la performance en la gestión de las organizaciones estudiadas. Es decir, de la observación de empresas que se encuentran alejadas de la zona de riesgo de quiebra, según determinan estos modelos, se pudo concluir que no siempre esta posición de solvencia, tiene un correlato con una gestión empresarial óptima o eficiente.

Por ese motivo se decidió realizar una extensión del proyecto de investigación original. En esta segunda etapa del proyecto, se espera poder estudiar, sobre la misma fuente masiva de datos financieros que se utilizó en la primera etapa, la pertinencia de modelos de evaluación de performance de gestión.

Del estudio bibliográfico que se realizó sobre técnicas vinculadas a la implementación de indicadores de gestión, se decidió centrar la segunda etapa del proyecto, en la evaluación de los denominados KPI (Key Performance Indicators - Indicadores Claves de Performance).

2. OBJETIVOS DE LA PRESENTE ETAPA DE DESARROLLO DEL PROYECTO DE INVESTIGACION.

- Evaluar el comportamiento de los modelos de evaluación de gestión empresarial basados en KPIs, cuando son aplicados a cantidades masivas de datos financieros, por medio de técnicas de Ciencia de Datos.
- En base a la evaluación realizada, proponer la discusión de los resultados observados, a fin de enriquecer los modelos analizados.
- Documentar las actividades realizadas, observando errores y aciertos, a fin de transmitir la experiencia obtenida en la implementación de las herramientas seleccionadas, en ambientes de cantidades masivas de datos y Ciencia de Datos.

3. REVISIÓN BIBLIOGRÁFICA.

Basaremos la revisión bibliográfica tomando en consideración los dos ejes temáticos centrales sobre los que se basa la investigación:

- KPI (Indicadores Claves de Performance).
- Big Data (Datos Masivos) y Data Science (Ciencia de Datos).

3.1. KPIs - Indicadores Claves de Performance. Visión estratégica.

En el contexto actual de una economía globalizada, altamente competitiva y sometida a una vorágine de cambios tecnológicos constantes, es casi imposible pensar en empresas que no guíen su accionar siguiendo un plan estratégico consistente.

¹ CAMPANARO, R. S., DIAZ, D. J., GARDENAL, L., MARCHESE, A. G. (2016): Análisis de estados contables aplicando XBRL y herramientas de inteligencia de negocios. DUTI 2016. Bahía Blanca. Argentina.

Para desarrollar e implementar su planeamiento estratégico las empresas cuentan con conocimiento adquirido, herramientas, modelos teóricos y compendios de buenas prácticas, siendo algunos de los más destacables:

- Definición de Misión, Visión y Valores.
- Balanced Scorecard (BSC).
- Mapas estratégicos.
- Modelos de Execution Premium.
- Estrategias de Océano Azul.

Todos estos recursos utilizados en la elaboración y monitoreo del plan estratégico de la empresa, en mayor o menor medida, se nutren de la información extraída de diferentes indicadores de gestión, dentro de los cuales podemos citar los KPIs.

Parmenter, D. define los KPIs como *“KPIs representan un conjunto de mediciones enfocadas en aquellos aspectos de la performance organizacional que son los más críticos para el actual y futuro éxito de la organización”*²

Si bien del análisis de los KPIs que utilizaremos en nuestra investigación, podemos concluir una relación directa de los mismos con el ámbito de la gestión operativa de la empresa, no podemos también dejar de mencionar su vinculación con el planeamiento estratégico de la misma.

Respecto a las características de esta vinculación (KPIs - Planeamiento Estratégico) hay diferentes enfoques entre autores.

Kaplan y Norton recalcan la necesidad de que exista una estrategia explícita para capturar los beneficios que arrojan los KPIs: *“KPI scorecards can drive improved operational performance, but unless they are accompanied by an explicit strategy to capture the benefits, the organization will experience disappointing outcomes”*³.

Parmenter D. sostiene que los KPIs tienen que tener una vinculación más relevante con los CSF (Critical Success Factors - Factores críticos de éxito) que con el Balanced Scorecard (Cuadro de Mando Integral) o los objetivos estratégicos de la empresa: *“For a performance measure to be a KPI it has to be linked to one or more of the organization’s critical success factors (CSFs), more than one balanced scorecard (BSC) perspective, and the organization’s strategic objectives.”*⁴

Este último autor plasma su visión de cómo se integran los KPIs con el plan estratégico de la empresa, en el marco de trabajo que se puede observar en siguiente gráfico:

² Traducción del autor: “KPIs represent a set of measures focusing on those aspects of organizational performance that are the most critical for the current and future success of the organization” Parmenter (2010): Key performance indicators: developing, implementing and using winning KPIs.

³ KAPLAN, R. S., NORTON, D. P. (2001): The strategy-focused organization: How balanced scorecard companies thrive in the new business environment.

⁴ PARMENTER, D. (2010): Key performance indicators: developing, implementing and using winning KPIs.

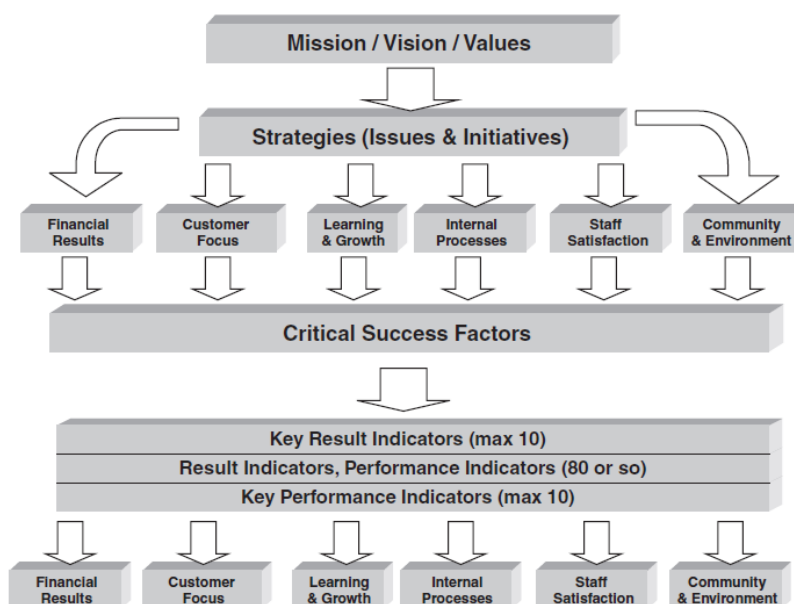


EXHIBIT 2.2 Journey from a Mission and Vision to Performance Measures that Work

Figura 1: viaje desde una Misión y Visión a las Medidas de Performance. Parmenter (2010).

En lo que respecta a la decisión de elección de KPIs, como indicadores de gestión a ser utilizados en nuestro proyecto, la misma se vio vinculada a la factibilidad de extracción en forma automática de datos financieros parametrizados de la fuente de datos que utilizaremos (Balances de empresas reportados a la SEC de USA, codificados en XBRL - eXtensible Business Reporting Language).

Algunos indicadores altamente difundidos en la academia y en el ámbito empresarial, tales como los derivados de implementar BSC (Balanced Scorecard - Cuadro de Mando Integral), nos presentaban el inconveniente de no poder obtener con facilidad variables de carácter cualitativo que usan en los mismos, ya que difícilmente las mismas, forman parte de los Estados Financieros de empresas.

3.2. Big Data (Datos Masivos) y Data Science (Ciencia de Datos).

Seguramente el aspecto diferencial de la investigación que llevamos adelante sobre análisis de indicadores de gestión, es la implementación de los mismos a fuentes masivas de Datos (Big Data) por medio de técnicas de inferencia de conocimiento centradas en Data Science (Ciencia de datos).

Hay coincidencia en caracterizar los desarrollos de Big Data, por medio de tres aspectos destacables conocidos como las 3 V: Volumen, Velocidad, Variedad.

Volumen: Cantidades masivas de datos, generalmente medidas en Gb, Tb, Pb.

Velocidad: Hace referencia a la disponibilidad de esa cantidad masiva de datos para poder ejecutar sobre los mismos búsquedas o procesos de análisis.

Variedad: Al tener primacía en la conformación de estos grandes repositorios de datos, información derivada de redes sociales, se debe contemplar la incorporación de datos estructurados, semi-estructurados y no estructurados.



Figura 2: tres ejes conceptuales de Big Data.

En lo que respecta a las técnicas y algoritmos de análisis de datos enrolados en lo que se define como Ciencia de Datos, nuestro enfoque estará guiado por aquellas que brindan la posibilidad de detección de patrones y relaciones de tipos insospechadas en las cantidades masivas de información financiera extraída de los Estados Financieros a analizar.

Tal como define Foster Provost, T.: *“ciencia de Datos envuelve principios, procesos y técnicas para entender fenómenos por medio de (automatización) análisis de datos”*⁵.

Dentro de las técnicas más relevantes del campo de Ciencia de Datos que analizaremos a fin de determinar su factibilidad y pertinencia para el logro inferencia de conocimiento que buscamos, podemos citar a **Data Management, Análisis Exploratorio, Visualización de datos, Aprendizaje de Máquinas (Machine Learning), Data Mining y Text Mining**

La mayoría de la literatura que abarca la problemática de Cantidades Masivas de Datos (Big Data) y su análisis por medio de técnicas de Ciencia de Datos (Data Science), coincide en identificar tres tipos de fuentes de datos masivos que se vienen analizando y que generan el impulso que esta disciplina viene teniendo en la actualidad:

- a) **IoT (Internet of Things - Internet de las cosas)**: refiere a los diferentes tipos de sensores, drones, imágenes satelitales, y en general todo tipo de dispositivos de medición automatizada.
- b) **Iniciativas de Open Government (Gobierno Abierto)**: en especial en lo que refiere a desarrollos de Datos Abiertos (Open-Data). Estas iniciativas son dirigidas por instituciones multinacionales como el BID (Banco Interamericano de desarrollo), cuyo propósito es que los gobiernos abran el acceso a la ciudadanía de las grandes bases de datos utilizadas en su gestión.
- c) **Redes Sociales**: las cuales generan día a día un constante flujo de información generalmente no-estructurada, relativo a diferentes aspectos de nuestras vidas, opiniones, sentimientos, etc...

De estas tres fuentes masivas de datos, la que mayor empuje ha dado al desarrollo de herramientas de bases de datos y de análisis predictivos de los mismos ha sido la tercera, las redes sociales.

Google ha facilitado la divulgación de desarrollos orientados a almacenar y gestionar cantidades masivas de datos no-estructurados. También ha fomentado el uso de la plataforma Apache-Hadoop. Hadoop es un desarrollo de base de datos NoSQL,

⁵ Traducción del autor. “Data science involves principles, processes, and techniques for understanding phenomena via the (automated) analysis of data”. FOSTER PROVOST, T. (2013): Data Science for Business.

basado en una infraestructura de almacenamiento distribuido, con tolerancia a fallos y basado en el algoritmo map-reduce, sobre datos altamente indexados.⁶

Existe también disposición a utilización por medio de licencia GNU- Open Source (uso gratuito) herramientas para la gestión y análisis de datos almacenados en plataforma Apache-Hadoop, tales como Spark⁷, Pig⁸, Hive⁹.

Los líderes de desarrollo de software mundial han incorporado los servicios de procesamiento de grandes cantidades de datos en la nube, facilitando el uso de sofisticadas herramientas de análisis de Ciencia de Datos.

Desarrollos como los de Google Cloud¹⁰, Microsoft Azure¹¹, o Oracle Cloud¹² son muestras de una tendencia irreversible, que abre las puertas a la concepción de un nuevo paradigma de gestión de información.

4. ASPECTOS METODOLÓGICOS.

El proyecto se basa en una investigación de tipo descriptiva, de características cuali - cuantitativas, cuyo principal enfoque se orienta al análisis de datos.

Otro aspecto que deseamos resaltar sobre la investigación a realizar es su carácter interdisciplinario. Debemos destacar que se debe trabajar con conocimientos vinculados a administración y con conocimientos vinculados al área de IT (Tecnología de la Información).

En este último caso, y específicamente dentro de la disciplina de Data Science (Ciencia de Datos), encontramos elementos ligados a Estadísticas (algoritmos utilizados), Administración (análisis del modelo de negocio sobre el cual se trabaja) y propios de IT (lenguajes de programación, librerías y componentes de software, etc.).

Respecto a las actividades relevantes planificadas, y a los desafíos que las mismas abren, podemos mencionar:

4.1. Perfil de IT (Tecnología de la Información).

- a) Proceso de ETL (Extracción, Transformación y Almacenamiento de datos). ANEXO 1.
- b) Análisis de almacenamiento de datos en Bases de Datos NoSQL.
- c) Modelado de consultas de datos orientadas a su análisis.
- d) Consulta inteligentes de datos.
- e) Depuración de datos.
- f) Implementación de técnicas de Data Science. ANEXO 2.
- g) Evaluación y depuración de resultados obtenidos.
- h) Análisis de conocimiento obtenido por aplicación de las técnicas de Data Science. Conclusiones.

4.2. Perfil de análisis financiero.

- a) Análisis y estudio del marco general de indicadores de gestión.
- b) Vinculación de indicadores de gestión con el planeamiento estratégico.
- c) Análisis y fichaje de bibliografía relevante sobre KPIs.

⁶ <http://hadoop.apache.org/>

⁷ <http://spark.apache.org>

⁸ <http://pig.apache.org/>

⁹ <http://hive.apache.org/>

¹⁰ <https://cloud.google.com/>

¹¹ <https://azure.microsoft.com/es-es/>

¹² <https://cloud.oracle.com/home>

- d) Selección de KPIs a utilizar.
- e) Filtrado de KPIs en función de extracción de datos de fuentes masivas de datos.
- f) Selección de la muestra de fuente de datos, en función de parámetros de análisis.
- g) Implementación de KPIs en fuentes masivas de datos.

5. RESULTADOS Y CONCLUSIONES.

El presente trabajo de investigación se encuentra en desarrollo. Al aplicar técnicas de data science se buscan relaciones de tipo insospechada, por lo cual el carácter de los resultados es impredecible en esta etapa de la investigación. Se espera como resultado potencial, la detección, vinculaciones entre variables relevantes, inferencia de patrones de comportamiento y descubrimiento de relaciones imprevisibles, que nos permitan ratificar la vigencia o postular nuevos modelos, tanto de predictibilidad de quiebras empresariales, como de evaluación de performance en la gestión. Esta última consideración nos lleva a la necesidad de planificar la metodología que se aplicará para evaluar los resultados obtenidos.

Cumplidas las tres primeras etapas mencionadas en el Perfil de análisis financiero, nos encontramos abocados a la depuración y selección de los KPIs relevados, tradicionalmente agrupados, tales como:

- Indicadores de Solvencia básica corriente
- Indicadores Financieros (Operativos)
- Indicadores de Liquidez y Flujo de Caja
- Indicadores de Valor Económico Añadido
- Indicadores de Eficiencia Financiera
- Indicadores desde la perspectiva de los clientes.

A la par de estos habituales indicadores, estamos contemplando el impacto que tienen las nuevas organizaciones, basadas en el conocimiento y donde su mayor valor se centra en el capital intelectual. Estas exigen otras maneras de evaluación, que se basará sobre una muestra de empresas a determinar en el futuro.

6. REFERENCIAS BIBLIOGRÁFICAS.

6.1. Área IT.

- BELL, J. (2015): Machine Learning: Hands-On for Developers and Technical Professionals. Willey.
- COLIN WARE, E. (2004): Information Visualization: Perception for Design. Morgan Kaufmann.
- FOSTER PROVOST, T. F. (2013): Data Science for Business. O'Reilly.
- HARRINGTON, P. (2012): Machine Learning in Action. Manning.
- WITTEN, IAN H., FRANK, E., HALL, M. (2011): Data Mining: Practical Machine Learning Tools and Techniques. Elsevier / Morgan Kaufmann.
- KUNCHEVA, L. I. (2014): Combining Pattern Classifiers. Second Edition ed. Wiley.
- TUFTE, E. R. (1990). Envisioning Information. Graphics Press.
- YU-WEI, CHIU (DAVID CHIU) (2015): Machine Learning with R. Cookbook. Pack Publishing.
- MORRISON, M. (2000): XML al descubierto. Prentice Hall. Madrid
- BERGERON, B. (2004): Essentials of XBRL: Financial reporting in the 21st century (Vol. 30). John Wiley & Sons.
- HOFFMAN, C., WATSON, L. (2009): XBRL for Dummies. John Wiley & Sons.
- MOHANTY, S., JAGADEESH, M., SRIVATSA, H. (2013): Big Data Imperatives. Apress.

FRAMPTON, M. (2014). Big Data Made Easy: A Working Guide to the Complete Hadoop Toolset. Apress.

STUBBS, E. (2014). Big Data, Big Innovation: Enabling Competitive Differentiation Through Business Analytics. John Wiley & Sons.

6.2. Área de Análisis Financiero.

WALSH, C. (2003): Key management ratios: master the management metrics that drive and control your business. Financial Times/Prentice Hall.

WALSH, C. (2006): Key management ratios: the clearest guide to the critical numbers that drive your business. Pearson Education.

PARMENTER, D. (2010): Key performance indicators: developing, implementing and using winning KPIs. John Wiley & Sons.

HUBBARD, D. W. (2017): How to measure anything: Finding the value of intangibles in business. John Wiley & Sons.

HUBBARD, D. W. (2009): The failure of risk management: Why it's broken and how to fix it. John Wiley & Sons.

BAKER, R. J. (2006): Measure What Matters to Customers: Using Key Predictive Indicators (KPIs). John Wiley & Sons.

CAMPANARO, R. S., DIAZ, D. J., GARDENAL, L., MARCHESE, A. G. (2016): Análisis de estados contables aplicando XBRL y herramientas de inteligencia de negocios. DUTI 2016. Bahía Blanca. Argentina.

KAPLAN, R. S., NORTON, D. P. (2001): The strategy-focused organization: How balanced scorecard companies thrive in the new business environment. Harvard Business Press.

MARR, B. (2015): Key Performance Indicators for DUMMIES. John Wiley & Sons.

ANEXO 1.

Consideraciones vinculadas a las actividades previstas - Perfil IT.

Fuente de datos:

La SEC (Securities and Exchange Commission) de USA viene recibiendo desde el 2005 reportes financieros de sus empresas reguladas, codificados con el lenguaje XBRL (eXtensible Business Reporting Language) es decir en formato apto para su explotación por sistemas computacionales. Esta fuente masiva de datos, al momento de desarrollarse el presente trabajo contaba con más de 160.000 reportes financieros de empresas.

Este repositorio de información económico-financiera representa una apreciable fuente de datos para el estudio de modelos analíticos y predictivos de comportamiento empresarial.

Proceso de ETL:

El estándar financiero-tecnológico utilizado por la SEC para que las empresas reguladas informen sus Estados Financieros, es XBRL.

XBRL es un sub-lenguaje derivado de XML (eXtensible Mark-Up Language), siendo este último, universalmente aceptado para la representación de información de todo tipo, para su uso en la red.

La representación de datos en XML se logra mediante el desarrollo de dos tipos de documentos que trabajan en conjunto. Un "schema", donde se define la tipología de datos y su esquema jerárquico de validación, y un "documento de instancia" donde se expresan los valores que adoptan los datos definidos en el esquema.

Una particularidad que se resalta en la implementación de XBRL es que el esquema de datos contemplado por XML se complejiza al requerir información estructurada de tipo financiera (Estados Contables), en donde tenemos que enfrentarnos con diversos tipos de relaciones entre datos (etiquetas, definiciones, cálculos, presentación, dimensiones, etc...). Estos esquemas complejos reciben el nombre de Taxonomías XBRL.

Lo mencionado en el párrafo anterior toma un realce significativo en el proceso de ETL (Extraction, Transform and Load - Extracción, Transformación y Almacenamiento), ya que no solamente se debe identificar inequívocamente las variables a extraerse de la fuente de datos, sino que se deben valorar los contextos definicional, dimensional y temporal de las mismas.

ANEXO 2.**Consideraciones vinculadas a las actividades previstas - Perfil IT.****Técnicas y algoritmos de Ciencia de Datos a ser evaluadas para determinar su pertinencia de implementación:**

Proceso de Análisis	Descripción	Algoritmos utilizados	Grado de dificultad
Data Management	Proceso previo al análisis de datos, donde se preparan los data sets para ser estudiados	NAs (Tratamiento de valores faltantes) Eliminación de ruido (noise reduction) Normalización Transformación de datos Detección de outliers	Bajo
Análisis Exploratorio	Técnicas estadísticas básicas al que se someten los datos para determinar relaciones, asociaciones, agrupamientos básicos y así obtener conclusiones preliminares en el análisis	Correlación de variables Intervalos de confianza Histogramas Densidad	Medio
Visualización de datos	Técnicas de comunicación visual de datos, basadas en la habilidad del cerebro humano de interpretar con mayor facilidad información presentada visualmente	Regresión Lineal Regresión local Scatter plots Box plots Cuartiles Gráficas de barras Gráficas de densidad	Medio
Aprendizaje de Máquinas (Machine Learning)	Técnicas que permiten a sistemas aprender de datos, por medio de un entrenamiento basado en resultados observados con anterioridad	Regresión Lineal Regresión Logística Least squares regression k-nearest Naïve Bayes K-means clustering Árboles de decisión Máquinas de vectores de soporte Random forests	Alto
Data Mining	Proceso no trivial de identificación de patrones válidos, novedosos,	Análisis de Cluster: K-means clustering K-medoids clustering	Alto

	potencialmente útiles y comprensibles en los datos (Witten, Frank, Hall, 2011)	Hierarchical clustering Estimaciones de densidad Expectation-maximization Reglas de Asociación Patrones: Eclap Apriori Sequence metrics	
Text Mining	Técnica especializada de Data Mining, enfocada a grandes cantidades de datos, en formato de texto, no estructurado, masivo y de mucha variabilidad	Procesamiento de textos Clustering de textos	Alto

CAMPANARO, R. S., DIAZ, D. J., GARDENAL, L., MARCHESE, A. G.. (2016): Análisis de estados contables aplicando XBRL y herramientas de inteligencia de negocios. DUTI 2016. Bahía Blanca. Argentina.