UNIVERSIDAD DE ANTIOQUIA

Apoyo a la dirección de servicios para la innovación de CIDET en la ejecución de proyectos de desarrollo e innovación y vigilancias tecnológicas

Autor
León Alexis Buitrago López

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica
Medellín, Colombia

2020

Apoyo a la dirección de servicios para la innovación de CIDET en la ejecución de proyectos de desarrollo e innovación y vigilancias tecnológicas

León Alexis Buitrago López

Informe de práctica
como requisito para optar al título de:
Ingeniero Electricista

Asesores

Nicolás Muñoz Galeano
Doctor en Electrónica de Potencia

Eduin García Arbeláez
Director (E) Dirección de Servicios para Innovación CIDET

Universidad de Antioquia
Facultad de Ingeniería, Departamento de Ingeniería Eléctrica
Ciudad, Colombia
2020

Apoyo a la dirección de servicios para la innovación de CIDET en la ejecución de proyectos de desarrollo e innovación y vigilancias tecnológicas

Resumen

El documento registra los servicios relacionados con la ejecución de proyectos de desarrollo, innovación y vigilancia tecnológica ejecutados para diferentes empresas del sector eléctrico, empresas conexas o en su defecto empresas de otros sectores a través de la dirección de servicios para la innovación de CIDET. Cada uno de los registros expone una descripción de la problemática de la empresa, descripción de la necesidad o solución que buscan, información más relevante para el mapa de ruta, herramientas o estrategias de búsqueda de información y resultados o solución propuesta. Finalmente, se documenta el desarrollo de la herramienta de gestión documental para el área de dirección de servicios para la innovación de CIDET.

Introducción

La tecnología, el mercado y el conocimiento son factores que actualmente cambian rápida y constantemente, por lo que obliga a las empresas a realizar inversiones en investigación y desarrollo e innovación. La investigación les permite adquirir conocimientos técnicos y científicos mientras que el desarrollo de esas tecnologías tiene como objetivo la fabricación de nuevos productos, procesos o materiales. Sin embargo, no solo se trata de adquirir conocimiento ya estructurado sino de perseguir la generación de nuevo. Por lo tanto, dentro de este proceso es fundamental que estén implícitos conceptos como la innovación y la creatividad (G.D., febrero de 2013 p.).

El objetivo de la implementación de la investigación y el desarrollo dentro de una empresa es lograr un proceso de innovación a través del cual se logre mejorar aspectos como: la calidad de los productos, servicio, proceso, la apertura de nuevos mercados, un incremento de ventas, una disminución en costos etc. De modo que, pueda lograrse una ventaja competitiva. No obstante, el éxito de la I+D para lograr la innovación solo se logra cuando se toman decisiones correctas y se conoce el entorno competitivo.

Existen metodologías que le permiten a las empresas tomar decisiones dejando de lado la intuición o la especulación y basándose más en esquemas de información. Esto disminuye la probabilidad de tomar decisiones erróneas lo que resulta en un incremento de la competitividad. La inteligencia competitiva y la vigilancia tecnológica son disciplinas a través de las cuales se pueden construir este tipo de metodologías. La primera está relacionada con el proceso de obtención, análisis, interpretación y difusión de información de valor estratégico sobre la industria y los competidores, que se transmite a los responsables en la toma de decisiones en el

momento oportuno (G.D., febrero de 2013 p.). El segundo concepto, por su parte consiste en la recolección de información especializada en un tema, para su posterior depuración y análisis, convirtiendo así la información en un conocimiento útil para la toma de decisiones estratégicas (Aguirre J.J., 2012).

En Colombia son pocos los casos de empresas o sectores que han adoptan procesos de I+D acompañados de herramientas de vigilancia tecnológica e inteligencia competitiva. Lo que se refleja en la dificultad en la toma de decisiones acertadas por parte de las organizaciones, la pérdida o el desaprovechamiento de oportunidades de negocio y el evidente atraso tecnológico en una gran parte de las grandes y medianas empresas. Por el contrario, en países como España existe una fuerte promoción y desarrollo en el campo de la vigilancia tecnológica e inteligencia competitiva a tal punto que se han desarrollado softwares como Vigiale, Hontza, Softtvt entre otros que facilitan este tipo de procesos y existe un gran porcentaje de empresas con áreas dedicadas solo a esta actividad. Se resalta, que existen normativas como UNE 166006:2006 AENOR que permiten estandarizar y hacerle seguimiento al proceso de vigilancia de modo que garantice su transparencia.

Objetivos

Objetivo General:

Apoyar a los profesionales de la dirección de servicios para la Innovación de CIDET en vigilancias tecnológicas e inteligencia competitiva enmarcados en la formulación, ejecución de proyectos de desarrollo tecnológico e innovación y que buscan contribuir a la construcción de mapas de ruta que permitan a las empresas del sector eléctrico u otros sectores en la toma decisiones y en el desarrollo de procesos de innovación o desarrollo tecnológico a través de los cuales puedan lograr una ventaja competitiva.

Objetivos Específicos:

- Apropiar las técnicas y herramienta para la realización de búsqueda en bases de datos especializadas y no especializadas.
- Depurar, sistematizar y analizar la información generado valor para los diferentes interesados
- Redactar informes para proyectos de desarrollo, innovación, vigilancia tecnológica e inteligencia competitiva.
- Participar en el desarrollo de una herramienta que permita la gestión del repositorio de información de la dirección de servicios para innovación de CIDET.
- Apoyar las actividades inherentes al desarrollo de la Corporación, como participación en eventos y otras áreas.

Marco Teórico

Para el desarrollo adecuado de vigilancias tecnológicas e inteligencia competitiva es importante conocer los conceptos relacionados con estas prácticas de forma que desde el punto de vista técnico se realice la actividad de una forma correcta y transparente. Por esto, a continuación, se hace un breve resumen de la terminología y la descripción de la metodología según la norma.

1. Vigilancia Tecnológica

La Vigilancia Tecnológica (VT) es una metodología aplicada para la obtención, análisis y difusión de forma especializada de la información, de forma que, se le da un valor agregado a esta con el propósito central de generar conocimientocomo insumo para latoma de decisiones estratégicas dentro de las organizaciones.

La vigilancia tecnológica, ante la necesidad de adquirir información de los entornos para la toma de decisiones, como metodología, amplía su campo de acción y no solo se enfoca en la información científica y técnica, sino que aborda otros marcos de acción, tales como: normativo, económico, comercial, competitivo, socio, cultural, ambiental, entre otros (Sanchez, – ALTEC 2009). En consecuencia, de esto, se han creado una clasificación según su ámbito de aplicación. Los principales tipos de vigilancia son:

• **Vigilancia competitiva:** se ocupará de la información sobre los competidores actuales y los potenciales (política de inversiones, entrada en nuevas actividades, entre otros).

• **Vigilancia comercial:** estudia los datos referentes a clientes y proveedores (evolución de las necesidades de los clientes, solvencia de los clientes, nuevos productos ofrecidos por los proveedores, entre otros).

• **Vigilancia tecnológica:** se ocupa de las tecnologías disponibles o que acaban de aparecer, capaces de intervenir en nuevos productos o procesos.

• **Vigilancia del entorno:** se ocupa de la detección de aquellos hechos exteriores que pueden condicionar el futuro, en áreas como la sociología, la política, el medio ambiente, la ingeniería, las reglamentaciones, etc. De manera general, las anteriores tipologías de vigilancia tecnológica que se han establecido para monitorear los entorno. Vale la pena indicar que todos los tipos de vigilancia se desarrollan por medio de la metodología del ciclo de vigilancia tecnológica, lo cual unifica los procesos para llevar a cabo todo tipo de estudio o informe.

2. Inteligencia Competitiva

El proceso de Inteligencia Competitiva (IC) convierte datos en información de referencia para que las organizaciones transformen esa información en conocimiento y éste a su vez en estrategias y acciones. En consecuencia, la

información inteligente tendrá la forma de "Alertas" sobre cambios importantes en el entorno que tienen implicaciones para la organización y sus planes y programas o la de "Propuestas de decisión" sobre ajustes que deban realizarse a programas, proyectos y metas que se encuentran en ejecución. Buscando la generación de información de valor agregado para el desarrollo de iniciativas de I+D+i y los procesos de planeación o toma de decisiones. La Inteligencia Competitiva es entonces un sistema organizacional de referenciación mediante el cual se confronta el direccionamiento y competencias de una organización o sector, identificando las tendencias económicas, sociales, tecnológicas, de mercado, de competencia y laborales, para generar desarrollo y crecimiento, basado en alertas de los vectores estratégicos del entorno competitivo, con el fin de que éstos anticipen sus respuestas a un entorno que es dinámico y cambiante (Aguirre J.J., 2012).

Según la literatura se hablan de cuatro tipos de inteligencia:

• **Económica y del entorno:** centra la observación sobre el conjunto de aspectos sociales, legales, medioambientales, culturales, que configuran el marco de la competencia.

• **Mercado:** dedica la atención sobre los clientes y proveedores de la cadena y mercados locales e internacionales.

• **Tecnológica:** centrada en el seguimiento de los avances del estado del arte tecnológico, y en particular, de la tecnología y de las oportunidades/ amenazas que genera.

• **Competitiva:** implica un análisis y seguimiento de los competidores actuales, potenciales y de aquellos con producto substitutivo.

3. Prospectiva

La prospectiva es una disciplina para el análisis de sistemas sociales que permite conocer mejor la situación presente, identificar tendencias futuras y analizar el impacto del desarrollo científico y tecnológico en la sociedad (Aguirre J.J., 2012). Con ello se facilita el encuentro de la oferta científica y tecnológica con las necesidades actuales y futuras de los mercados y de la sociedad. La prospectiva no es predicción, utopía, ciencia ficción, profecía ni adivinación. La VT y la IC se complementan con la prospectiva. Para construir escenarios prospectivos sobre un tema o sector, se requiere, previamente, adelantar un proceso de vigilancia tecnológica o inteligencia competitiva que proporcione contexto y realidad acerca del tema en estudio, de modo tal, que la prospectiva tenga este insumo como punto de referencia y partida para prospectar el futuro por medio de escenarios (Aguirre J.J., 2012).

4. Aspectos Normativos

Una de las normas más representativa a nivel internacional que certifica los procesos de VT/IC fue desarrollada en España por el Comité Técnico de Normalización AEN/CTN 166 "I+D+i" donde especifica que la norma UNE 166006 Ex: 2006 AENOR parametriza un "Sistema de Vigilancia Tecnológica" que está orientada a empresas, organismos de apoyo a la I+D+i y a proveedores de VT. Esta norma está alineada con otras normas de sistemas de gestión como pueden serla UNE-EN ISO 9001:2000 y la UNE-EN ISO 14001:2004, y en especial con la UNE 166002:2002. La finalidad es aumentar la compatibilidad con dichas normas en beneficio de la comunidad de usuarios y permitir a las organizaciones alinear su propio sistema de gestión con los de estos otros sistemas (Aldasoro Alustiza, July 18-20, 2012.). En la tabla 1 se muestra un resumen de las normas que aplican a cada disciplina.

Norma	Vigilancia tecnológica	Inteligencia competitiva
Une 166000:2006	X	
Une 166001:2006		
Une 166002:2006	X	
Une 166005:2004 IN	X	
Une 16606:2011	X	X
Une 166007:2010 IN	X	

Tabla 1. Referencias a la vigilancia tecnológica e inteligencia competitiva en las normas de serie 166006 de AENOR (Aldasoro Alustiza, July 18-20, 2012.).

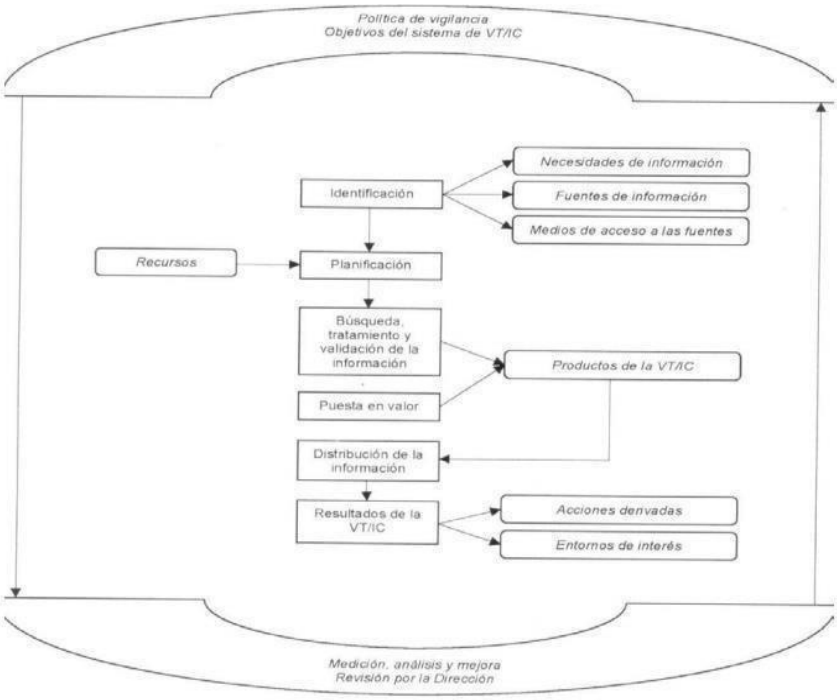


Figura 1. Proceso de VT/IC según la norma UNE-EN 166006:2011 *(Aldasoro Alustiza, July 18-20, 2012.)*.

El proceso de vigilancia tecnológica e Inteligencia competitiva debe cumplir un conjunto de pasos a través de cuales se garantiza que las organizaciones puedan implantar modelos de manera efectiva evitando sobrecostes económicos y de tiempo innecesarios. Esto además posibilita que las organizaciones elaboren su propio sistema de gestión mediante la aplicación de las normas AENOR.

Metodología

La metodología de trabajo se va a centrar en registrar los servicios de vigilancia tecnológica e inteligencia competitiva que se apoyan, estructurando la información de acuerdo con el esquema propuesto en la figura 2 y definido por la norma.

Figura 2. Vigilancia tecnológica de acuerdo con la norma UNE 166006: 2006 AENOR.

El reporte por cada uno de los servicios tendrá la siguiente estructura:

El primer paso de acuerdo con la figura 2 es la identificación de las necesidades que tiene la empresa. Dentro de esta etapa entonces se mencionarán los siguientes aspectos:

- Descripción general de la problemática que las empresas exponen teniendo en cuenta aspectos como: sector o área trabajo, situación actual, posible solución.

- Atendiendo a las preguntas del mapa de empatía para el cliente: ¿qué piensa y siente?, ¿qué ve?, ¿qué dice y hace?, ¿qué oye?, Obstáculos o esfuerzos, resultados deseados, necesidades.

El segundo paso es la planeación para esta etapa se van a definir los recursos para la vigilancia, es decir, se va a definir si se va a hacer búsqueda de patentes, artículos científicos, información especializada, búsqueda de mercado, estadística, información de base de datos no estructuradas etc. O una combinación de estos recursos de acuerdo con las necesidades de la empresa. Además, se definirá la estructura y periodicidad de presentación de resultados con el fin de mostrar el progreso en la vigilancia de modo que se pueda recibir retro alimentación y dirigir de la mejor forma el proceso haciendo que sea más enriquecedor.

En la tercera etapa se realiza una recolección y almacenamiento de la información dejando evidencia de la búsqueda a través de un cuadro que contenga las palabras cable y las referencias de donde se obtuvo la información.

En la cuarta etapa se realiza un análisis de la información recolectada de modo que se puedan extraer conclusiones a partir de las cuales se pueda crear un mapa de ruta a través del cual se den soluciones, recomendaciones y herramientas para la correcta toma de decisiones.

En la quinta etapa se hará una descripción del aprendizaje y/o conclusiones generales de cada uno de los servicios de vigilancia o inteligencia competitiva en los que se participa resaltando las habilidades y competencias que se desarrollan para la vida profesional.

Finalmente se describirá el proceso de desarrollo de la herramienta de gestión para el repositorio de información de la dirección de servicios para la innovación de CIDET. La descripción se centrará en: Descripción del problema, descripción de la solución y descripción de la herramienta.

Resultados y análisis

En las siguientes secciones se mostrarán los aspectos más relevantes y resultados de los diferentes proyectos desarrollados por la unidad de vigilancia tecnología de la dirección de servicio para la innovación CIDET, a través de estos proyectos podrá visualizar de forma general varios aspectos del sector eléctrico y otros sectores en la actualidad. Algunos de estos aspectos son: desarrollo tecnológico, problemáticas, tendencias, el mercado, aplicación de técnicas de ciencia de datos etc.

1. Proyecto: Apropiación de herramientas de vigilancia tecnológica.

La metodología utilizada en la ejecución de los servicios de prospectiva tecnológica y vigilancia (tecnológica y competitiva) así como de la gestión del conocimiento asociado a la innovación de acuerdo con las necesidades específicas de la Empresa; dicha metodología es el resultado de la adaptación de la guía de la norma UNE 166006:2018 para sistemas de vigilancia tecnológica e inteligencia competitiva, en la cual CIDET se encuentra certificado, y de diferentes experiencias internacionales recopiladas por la Agencia Internacional de Energía (IEA) en la edición del año 2014 del documento: "Energy Technology Roadmaps, a guide to development and implementation", que tiene por objetivo proveer a países y compañías el contexto, información y herramientas necesarias para diseñar, manejar, e implementar mapas de ruta que correspondan a sus propias circunstancias y objetivos.

En la Figura 3 se presenta el diagrama metodológico propuesto, el cual comprende dos tipos de actividades complementarias. En la parte superior del diagrama aparecen las actividades asociadas a la prospectiva tecnológica y en la parte inferior las asociadas a la vigilancia (tecnológica y competitiva). Dichas actividades se desarrollan en cuatro (4) fases: Planeamiento y visualización; alcance y estrategia de ejecución; ejecución del servicio y validación de resultados con expertos; y socialización de resultados finales, y son aplicadas de acuerdo con las características y requerimientos de cada solicitud de servicio.

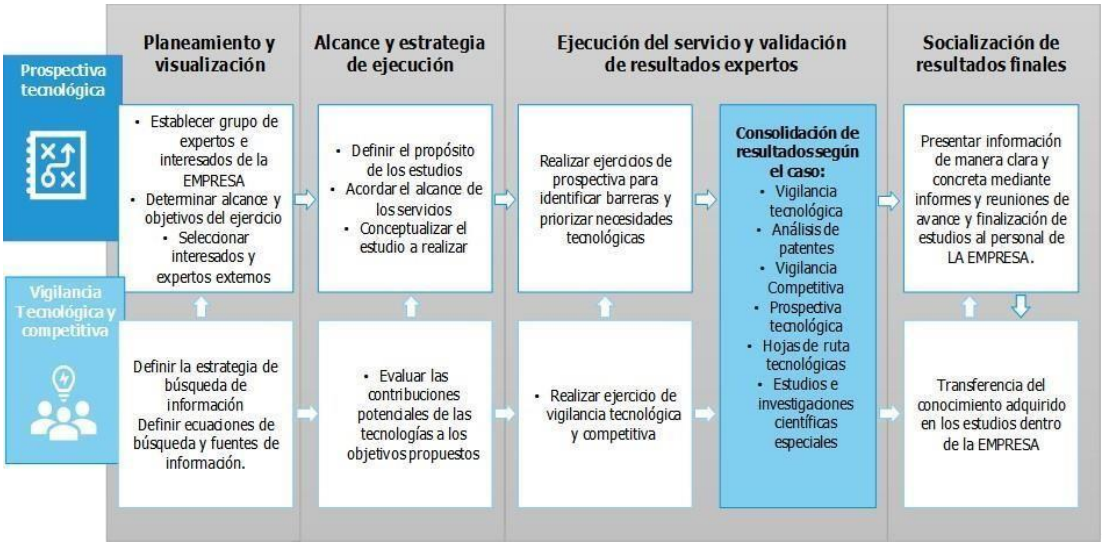


Figura 3. Esquema General para ejecución de los servicios - UNE 166006:2018 e IEA

En cuanto a las actividades asociadas a la prospectiva tecnológica, estas recogen la participación de expertos multidisciplinares en tecnología, política, economía, finanzas, ciencias sociales y otras disciplinas para formular metas e hitos del mapa de ruta, identificar brechas, determinar prioridades y asignar tareas. El juicio de expertos también es necesario para elegir entre posibles escenarios o las opciones reveladas por las actividades de análisis de información. Los objetivos de estas actividades son:

- Construir consensos, objetivos y metas
- Evaluar y verificar supuestos (costo de tecnología, métricas de desempeño)
- Identificar aspectos técnicos y barreras regulatorias
- Definir transformaciones tecnológicas para superar barreras
- Desarrollar estrategias de implementación y priorizar alternativas

La construcción de los resultados a partir de consensos con los expertos, tanto internos como expertos, ayuda a considerar varios puntos de vista e identificar escenarios de apropiación tecnológica. Este enfoque, facilita la transferencia de conocimiento ya que los principales interesados ya están familiarizados con estos.

El juicio de los expertos es soportado por las actividades de análisis de información a partir de los ejercicios de vigilancia y la revisión de experiencias, identificación de casos de éxito y de fracaso, que permiten un aprendizaje de relevancia para evitar repetir errores observados anteriormente.

Por otra parte, las actividades asociadas a la vigilancia tecnológica e inteligencia competitiva se enfocan en recopilar, procesar y analizar la información que permita orientar la toma de decisiones y los servicios de prospectiva, permitiendo identificar tendencias tecnológicas con impacto en el negocio de la compañía, cambios políticos y regulatorios, estrategias de los competidores entre otros elementos.

A continuación, se describen en detalle los elementos principales de la metodología propuesta, a partir de los procesos de vigilancia tecnológica e inteligencia competitiva, la prospectiva tecnológica y las herramientas utilizadas para su ejecución.

Vigilancia tecnológica e inteligencia competitiva

La Vigilancia Tecnológica es un proceso organizado, selectivo y permanente, basado en la captura de información sobre ciencia y tecnología en un sector de interés determinado; posteriormente, esta es seleccionada, analizada, difundida y comunicada a los decisores para convertirla en los conocimientos necesarios y suficientes para tomar decisiones inteligentes, minimizando riesgos y anticipándose a los cambios.

En la Figura 4, se muestra el diagrama de flujo para la elaboración de la vigilancia tecnológica e inteligencia competitiva adaptado de la norma UNE 166006:2018, el cual representa un esquema iterativo entre la definición del alcance y los entregables en busca de alcanzar los objetivos del estudio.

Figura 4. Diagrama de flujo de una vigilancia tecnológica e inteligencia competitiva - UNE 166006:2018.

Vale la pena resaltar que el proceso de vigilancia tecnológica e inteligencia competitiva ejecutado por CIDET, no es espionaje ni cuenta con herramientas o

prácticas para la obtención de información reservada o por fuera de los límites legales. La vigilancia se basa en la captación, análisis, síntesis, y utilización de la información pública existente. Su correcta interpretación y difusión impulsan la capacidad de claridad y anticipación de la empresa, sin necesidad de recurrir a prácticas poco éticas de obtención de información sobre competidores, estrategias, entre otras. A continuación, se describen brevemente cada uno de los pasos estipulados en la metodología.

- **Definición del alcance:** Partiendo de la identificación de necesidades puntuales y de una contextualización inicial del campo, desarrollo y/o actividades de la empresa a la cual se le prestará el servicio de vigilanciatecnológica.
- **Captura de información:** Una vez se define el alcance del estudio, los expertos pertenecientes al equipo de trabajo de CIDET, de acuerdo con su experiencia, detallan las estrategias de búsqueda. Como fuentes de información se usan recursos de bases de datos estructuradas (patentes, artículos científicos, memorias de conferencias, documentos de trabajo) y fuentes de información no estructurada (páginas web, fabricantes y compañías). Dicha información es validada por medio de consulta a expertos en la temática de interés en conjunto con ejercicios de recolección y análisis de información primaria según sea el caso. En esta etapa se encuentran unas subactividades que direccionan la manipulación de la información:

    - **Definir ecuaciones y estrategias de búsqueda:** Se plantean las estrategias de búsqueda, definidas en conjunto con los expertos del cliente y se construyen las ecuaciones de búsqueda que se usarán en las bases de datos estructuradas de acuerdo con esa información primaria recolectada.

    - **Organización y sistematización de información:** Se realiza la búsqueda de información del tema requerido. En esta etapa se identifican y validan las fuentes, palabras claves, subtemas y criterios de selección. Una vez realizadas las búsquedas, se da a conocer el detalle del proceso de búsqueda de información por medio de las ecuaciones registradas en la bitácora y se procede con la depuración de la información por medio de una herramienta especializada en el análisis de grandes volúmenes de información. En la bitácora se registra: fecha de la búsqueda, palabras clave, ecuación con conectores boleanos, base de datos en la que se realiza la búsqueda, filtros aplicados (ej. clasificación IPC o CPC), cantidad de hallazgos y enlace que direcciona elhallazgo.

    - **Identificación de principales documentos de interés:** A partir de la información recopilada y depurada, se identifica información relevante

acerca del objeto y alcance propuesto. Estos son los documentos de los cuales se empieza a hacer un análisis más profundo, de donde se extrae la información más importante y los que se van a almacenar según sea el caso.

- **Análisis bibliométrico:** Durante la búsqueda de información, es importante realizar un análisis bibliométrico para identificar tendencias de los investigadores, países, empresas, y demás involucrados que usen o provean tecnologías para el sector de estudio específico a nivel mundial. También, identificar quiénes lideran la investigación para contemplar posibles alianzas con el sector académico y estar al tanto de avances y prospectivas en el tema.

- **Depuración, Selección y Análisis:** Luego de capturar la información y tener a disposición una primera fuente base, se procesa por medio de análisis por parte del equipo de CIDET y se depura pasando por filtros a criterio de los analizadores los cuales, agregan valor a dicha información seleccionando aquella que es útil y definiendo las primeras conclusiones. En esta fase, la información relevante es adquirida y convertida en conocimiento que posteriormente se contrastará de acuerdo con los lineamientos de la empresa. En esta etapa los datos son procesados en un software especializado en minería de datos (ej. VantagePoint), con el fin de clasificar, depurar e identificar información relevante para el resto de la actividad.

- **Difusión de resultados:** Del previo análisis a la información obtenida, se elabora una serie de documentos (informes, presentaciones, hojas de cálculo, mapas de ruta, infográficos, entre otros) de acuerdo con las solicitudes del cliente y con el fin de apoyar asertivamente la toma de decisiones estratégicas. Adicional, estos resultados se van registrando para tener a disposición la trazabilidad del ejercicio e ir alimentando un repositorio o banco de información en la cual el cliente podrá acceder al documento original si desea profundizar en algún hallazgo.

- **Realimentación y socialización:** Luego de tener los documentos con la información de interés convertida en conocimiento, se hacen una serie de socializaciones donde se obtiene retroalimentación con el cliente a fin de validar dicha búsqueda y redireccionar o continuar en el proceso. Estas sesiones pueden ser presenciales, virtuales, foros o boletines dependiendo de cómo se organice directamente con las partes interesadas. De aquí se obtienen los insumos para continuar con el proceso iterativo, de ser necesario, o concluir los hallazgos.

- **Mapeo de oportunidades de Ciencia, Tecnología e Innovación:** CIDET a través de un proceso cíclico y periódico mapea oportunidades para el desarrollo de proyectos de Ciencia, Tecnología, Investigación y emprendimiento, utilizando para ello herramientas como Copernic Agent Basic (identifica cambios en páginas web previamente indexadas y envía información sobre cambios en ellas), servicios de alertas y búsquedas en la web a través de una base de datos predefinida.

- **Verificación cumplimiento de objetivos:** Dentro de las reuniones periódicas de seguimiento, se obtienen los insumos para continuar con el proceso iterativo, de ser necesario, o concluir los hallazgos. Es una etapa decisiva para asegurar el cumplimiento de los objetivos planteados.

- **Entregables:** Finalmente, después de completar las iteraciones necesarias, se hace entrega del proceso de vigilancia tecnológica con toda la documentación registrada y almacenada a lo largo de los procesos iterativos que responde a los objetivos planteados; esto, como insumo para la empresa y a la cual tiene disposición, incluso, de los documentos originales de los cuales se hizo extracción de la información relevante a fin de que pueda profundizar si lo considera necesario.

En la Tabla 2, se pueden observar las herramientas utilizadas durante la ejecución de los servicios de vigilancia tecnológica, la pertinencia del tipo de herramienta y la estrategia de búsqueda será tenida en cuenta en la discusión del alcance de cada necesidad en particular.

Tipo	Herramientas
Fuentes de información artículos científicos	• IEEE Xplore • e-CIGRE • ScienceDirect • Scopus • Scielo • Redalyc • Web of science
Bases de datos de documentos de patente	• Esp@cenet • Latipat • Patent Scope • Invenes • SIC • Google Patent • AcclaimIP
Herramientas y software de análisis inteligente (Metabuscadores)	• Observa • Copernic • Intelligo • CiteSeerX • Recolecta • Carrot2

Tipo	Herramientas
Herramientas especializadas en el análisis de grandes volúmenes de información (Minería de datos)	• VantagePoint
Gestores bibliográficos	• Zotero • Mendeley

Tabla 2. Síntesis *de herramientas vigilancia tecnológica*

De igual forma, de acuerdo con las características y requerimiento de cada servicio se identificarán y desarrollarán los factores críticos de vigilancia, entendidos como los factores externos a la organización que afectan de modo crítico a su competitividad y que deben ser vigilados. Algunos de los factores que se tendrán en cuenta son los siguientes:

- Tecnologías que están emergiendo y representan retos u oportunidades para la empresa.
- Tecnologías se están abandonando por estar maduras o no estar triunfandoen el mercado.
- Cambios políticos y regulatorios.
- Oportunidades no ocupadas por otras empresas.
- Movimientos estratégicos de los competidores: alianzas, compras, etc.
- Nuevos productos.
- Nuevos países de exportación (oportunidades de negocio/amenazas)
- Nuevas tendencias del mercado (necesidades)
- Nuevos y actuales proveedores
- Nuevas normas técnicas.

Prospectiva tecnológica

La prospectiva tecnológica es un proceso sistemático que analiza el estado actual y las perspectivas de progreso científico y tecnológico para identificar áreas estratégicas de investigación y tendencias tecnológicas en las que concentrar los esfuerzos de inversión y así obtener los mayores beneficios para la Empresa. La prospectiva tecnológica está orientadas a un conjunto de técnicas que permiten definir la relevancia de una tecnología en un momento futuro. Una característica principal de la prospectiva es que parte de la existencia de varios posibles futuros los cuales se enmarcan en un contexto dado, que puede ser bajo la jurisdicción de un país, un sector o una empresa.

La finalidad de la prospectiva tecnológica esfacilitar la toma de decisiones donde la tecnología constituye un factor cada vez más determinante, y en el que el propio ritmo de cambio tecnológico, cada día más acelerado, incorpora un grado creciente de incertidumbre. Por tanto, en este tipo de ejercicios toma gran

relevancia el concepto de expertos temáticos y herramientas que permitan construir los escenarios futuros a partir de la información disponible en el presente.

El proceso general para realizar los servicios de prospectiva tecnológica se puede observar en la Figura 5, el cual tiene en cuenta dos insumos principales para la ejecución, la línea base e información relevante para la construcción de futuros y el conocimiento de expertos temáticos internos y externos a la empresa. Dichos insumos se utilizan para alimentar las herramientas de prospectiva permitiendo la construcción de diferentes futuros que son objeto de análisis para construir recomendaciones y lineamientos estratégicos.

Figura 5. Esquema proceso de prospectiva tecnológica.

En la Tabla 3 se puede observar un conjunto de métodos y herramientas para realizar prospectiva tecnológica, la pertinencia de y aplicación de cada uno de ellos en los servicios será definida de acuerdo con los objetivos y el alcance.

Tabla 3. Síntesis de métodos y herramientas prospectiva tecnológica.

Tipo	Métodos y herramientas
Métodos basados en el conocimiento de los expertos para desarrollar visiones y escenarios a largo plazo.	• Grupos de expertos. • Brainstorming. • Mapas mentales. • Talleres de análisis de escenarios. • Método Delphi. • Análisis de impactos cruzados.
Métodos para identificar puntos de acción clave para determinar estrategias de planificación.	• Análisis Estructural. • Ábaco de Régnier • Análisis SWOT. • Tecnologías críticas / clave. • Árboles de relevancia. • Análisis morfológico. • Modelos de madurez tecnológico
Métodos cuantitativos (basados en hipótesis) que utilizan estadísticas y otros datos para realizar predicciones.	• Extrapolación de tendencias. • Modelos de simulación y dinámica de sistemas. • Mapas de ruta/hojas de ruta tecnológica

El proceso para la definición de una metodología a seguir se basa en la necesidad de tener un plan energético soportado en dos pilares: la convocatoria a expertos sobre el sector energético y desde allí construir una prospectiva que le permita tener un plan de desarrollo.

Esta dialéctica de proceso se desarrolla en las siguientes etapas de proceso:

- Diseño del problema
- Exploración del problema
- Priorización sistema
- Explicación del sistema
- Construcción de escenarios
- Simulación dinámica
- Plan de Desarrollo
- Evolución

El ejercicio inicia por una exploración de otros estudios y el levantamiento de información para el sector de interés en el ámbito mundial, a partir de ejercicios de vigilancia tecnológica e inteligencia competitiva. Con esa información se busca constituir una base de conocimiento para orientar las decisiones de los expertos y alimentar posibles modelos de simulación de futuro, en caso de que se considere necesario.

La siguiente fase consiste la aplicación de métodos basados en el conocimiento de los expertos para desarrollar visiones y escenarios a largo plazo. Por ejemplo, es posible aplicar la consulta Delphi, como mecanismo anónimo para conocer los principales drivers de transformación tecnológica y los impactos en el negocio para la empresa, ésta es planeada para tener una cobertura amplia de expertos internos y externos. Los resultados de la primera ronda Delphi son la base para la definición y realización de la encuesta de la segunda y más rondas.

Otras alternativas consisten en validar el estado de madurez de las tecnologías mediante métodos como Ciclo de Vida, Hype Cycle de Gartner o TRL (Technology Readiness Level, por sus siglas en inglés). Esta validación nos permitirá tener claro que tanto está preparada una tecnología para entrar al mercado. Tanto esta información, como la del juicio de expertos, son insumos para la construcción de mapas de ruta, que, alineados con la estrategia de la empresa, permiten plantear en un horizonte de tiempo la implementación de nuevas tecnologías, emprendimientos, y nuevas estrategias de mejoramiento.

Los resultados más importantes son revisados mediante una discusión directa para confrontar las opiniones de los participantes con las del equipo de CIDET, mediante métodos para identificar puntos de acción clave y determinar estrategias de planificación. En tal caso, se realiza un taller utilizando la metodología Ábaco de Régnier. En dicha discusión se busca validar los resultados Delphi de manera directa y consensuada, generando discusión y argumentación de posiciones, que pueden llevar a retractaciones o cambios de posición de los diferentes expertos.

Los resultados validados son el insumo para el ejercicio de análisis estructural, en el cual se consulta a expertos, lográndose obtener las áreas estratégicas o claves en la explicación del sistema. La ubicación de las diferentes áreas en el mapa estratégico permite seleccionar las áreas temáticas preponderantes, alrededor de las cuales individuos y organizaciones toman decisiones y definen sus retos y proyectos.

Mientras se llevaba a cabo el ejercicio prospectivo es posible utilizar herramientas cuantitativas para validar las hipótesis de los expertos, como por ejemplo herramientas de dinámica de sistemas, en todo caso estas herramientas serán validadas y aprobadas por los expertos de la Empresa. Al final del ejercicio prospectivo se puede ver el comportamiento de algunas variables importantes asociadas a la dinámica de las tecnologías, teniendo en cuenta acciones de apuesta acordes con la construcción de un escenario deseado, como es la implementación de estrategias y políticas de interés para el desarrollo del sector en relación con las áreas temáticas determinadas como relevantes a lo largo del ejercicio.

Herramientas de análisis.

Los procesos de búsqueda y evaluación de una tecnología o temática se pueden realizar a través de diferentes tipos de bases de datos, las diferentes clasificaciones son:

Fuentes de datos estructurados: Son bases de datos almacenadas de una forma sistemática donde el objetivo final será facilitar su posterior utilización en la búsqueda de un tema en específico. Entre las más reconocidas cabe mencionar bases de datos como Scopus, ScienceDirect, Oficinas de patentes, entre otras.

Fuentes de datos no estructurados: Se describe de forma genérica a los datos que no están contenidos en una base de datos o algún otro tipo de estructura de datos. Los datos no estructurados pueden ser textuales o no textuales. Se pueden generar en mensajes de correo electrónico, presentaciones PowerPoint, documentos de Word, software de colaboración y mensajes instantáneos. La información no estructurada de carácter no textual proviene de medios como imágenes JPEG, archivos de audio MP3 y archivos de vídeo Flash.

Análisis de madurez tecnológica: Son técnicas o herramientas que permiten evaluar la preparación de tecnologías para llegar al mercado y ser apropiadas por éste. Para evaluar la madurez de las tecnologías CIDET se apoya en las siguientes técnicas:

- **Ciclo de vida de la tecnología o Curva en S:** creada por Richard N. Foster (1987) relaciona el esfuerzo que se debe realizar, medido en recursos utilizados, para desarrollar una tecnología. Indica la evolución de la tecnología en el tiempo, donde a medida que aumenta su nivel de madurez hay que hacer mayores

esfuerzos para aumentar el rendimiento esperado de la tecnología (ver Figura 6).

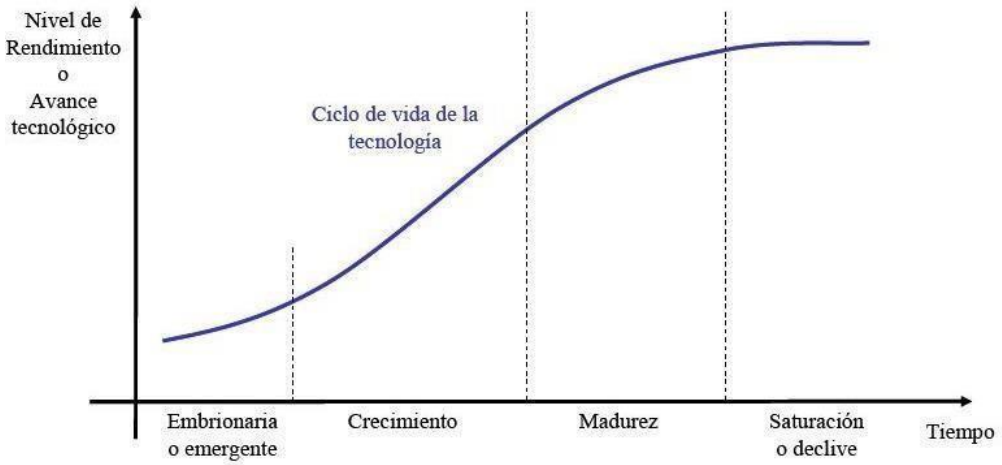


Figura 6. Curva ciclo de vida de la tecnología o curva en S.

- **TRL:** El nivel TRL es una técnica desarrollada por la NASA en los años 70 para evaluar la madurez de una tecnología previa a la integración de esta tecnología en un sistema (ver Figura 7).

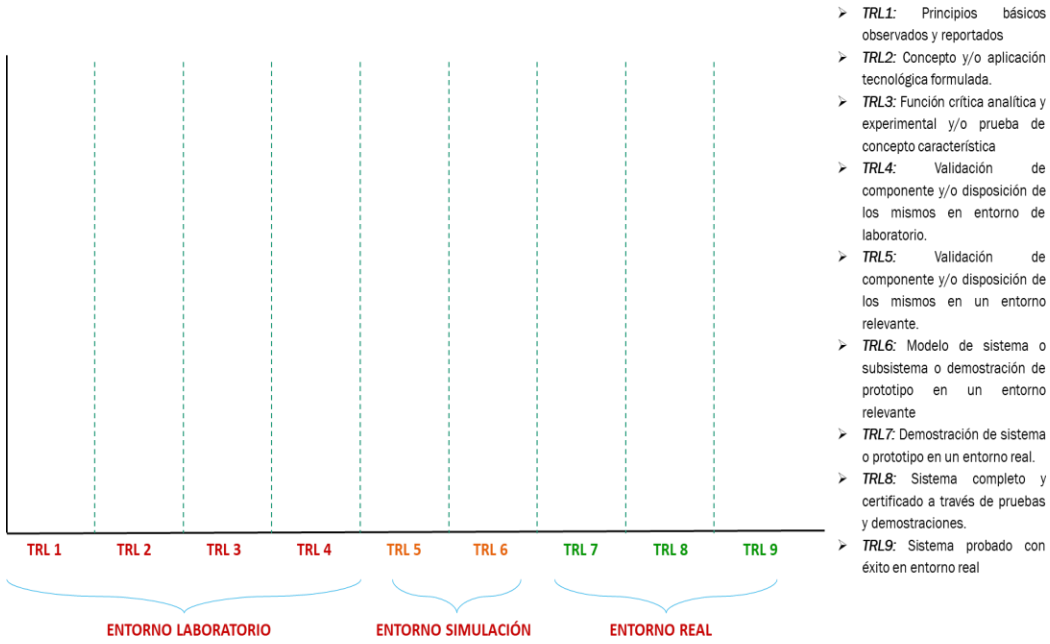


Figura 7. Esquema de TRL para tecnologías.

- **Hype Cycle – Gartner:** es una representación gráfica de la madurez, adopción y aplicación comercial de una tecnología específica (ver Figura 8).

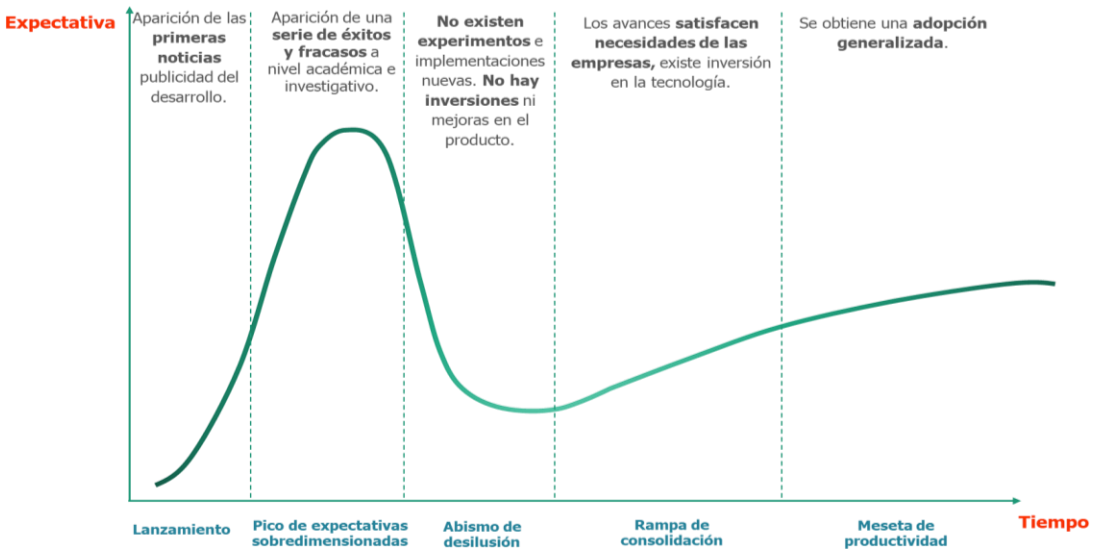


Figura 8. Hype Cycle – Gartner.

Se resalta que la calificación entre métodos de análisis de madurez de la tecnología es transferible, esto significa que la valoración hecha a través de los TRL puede visualizar a través de la Curva de Gartner, la selección entre un método u otro depende del vigía y de las características que más se quieran resaltar de la evaluación tecnológica.

2. Proyecto: Desarrollo de una herramienta para el diagnóstico decapacidades empresariales para la ejecución de proyectos de generación deenergía.

CIDET en su papel como promotor de desarrollo para el sector eléctrico participa de proyectos gubernamentales a través de los cuales se quiere incentivar la participación de las empresas del sector en proyectos jalonados por la innovación y de impacto social. Actualmente esta labor de promoción que CIDET realiza se lleva a cabo en compañía con INNPULSA que es la agencia de emprendimiento e innovación del Gobierno Nacional que acompaña la aceleración de emprendimientos de alto potencial y los procesos innovadores a través de apoyo económico para las empresas que participan de los proyectos. Uno de los intereses de esta entidad es llevar el fluido eléctrico a las zonas no interconectadas Colombia, con el objetivo de generar un desarrollo económico, equidad y oportunidades para todos los colombianos. El proyecto E2, por ejemplo, busca llevar energía eléctrica a la Guajira a través de instalación de paneles solares en zonas vulnerables. El papel de CIDET en este proyecto en particular se basa en el acompañamiento en el fortalecimiento de las capacidades empresariales de las empresas que participaran en la ejecución del proyecto. Para este fortalecimiento CIDET ha desarrollado una estrategia de diagnóstico y evaluación para las

empresas, para este proceso se desarrollo una herramienta en Excel con la que se busca facilitar inicialmente el proceso de valoración de las empresas y posteriormente facilitar el diagnóstico de estas. Con la herramienta se busca además brindar a las mismas empresas "una foto" del nivel de sus cualidades empresariales para el desarrollo del proyecto, y dará luz a cuáles son los aspectos que se deben mejorar. Luego, esto permitirá a CIDET identificar la línea base en cuanto a capacidades y madurez de la solución energética propuesta. Este diagnóstico, guarda coherencia con las dimensiones del autodiagnóstico incluido en la convocatoria FNCER-ZNI 2019, y servirá como insumo para el panel de evaluación, respecto de la selección del proyecto piloto.

Las empresas beneficiarias serán evaluadas teniendo en cuenta las dimensiones indicadas en la Figura 9: Dimensiones herramienta de diagnóstico.

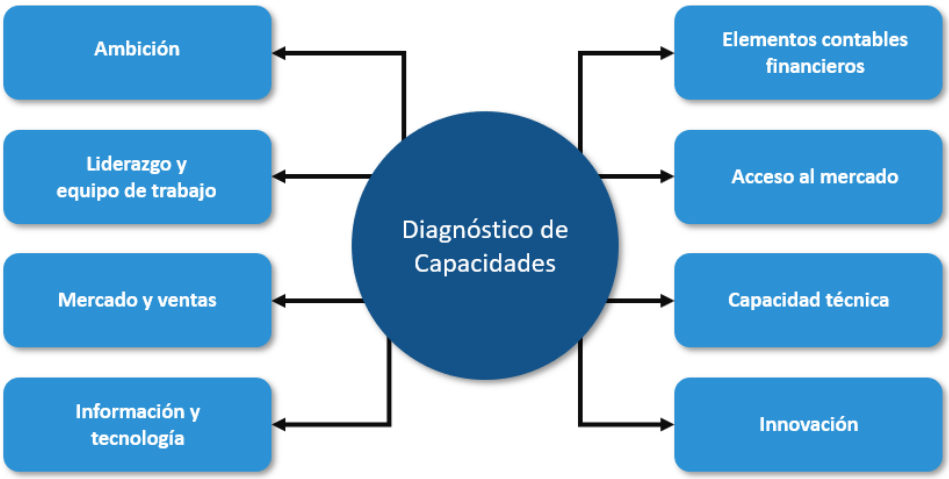


Figura 9. Dimensiones herramienta de diagnóstico

La meta final de esta consultoría consiste en fortalecer las capacidades de las empresas y sus propuestas, con miras a un escalamiento de sus negocios y de sus esquemas de electrificación rural. Por esta razón, el objetivo principal de la herramienta de diagnóstico es establecer una línea base como insumo para el diseño de los planes de fortalecimiento y escalamiento con cada una de las empresas beneficiarias. Esta herramienta es de fundamental importancia, pues a partir de los resultados del diagnóstico se establecerán todas las acciones futuras de la consultoría.

La aplicación de la herramienta también busca otros propósitos, como:

- Obtener información de contexto de cada empresa, que permita enfocar mejor las orientaciones que CIDET emitirá para cada una.
- Identificar oportunidades de escalamiento más recomendables y eficaces, con base en la naturaleza y orientación de mercado de cadaempresa.

- Establecer un nivel de "madurez comparable" inicial para las empresas seleccionadas, como insumo para el diseño de la herramienta de evaluación de propuestas para INNPULSA.

- Establecer puntos de control para la medición del progreso de las empresas seleccionadas, que permitan evaluar la efectividad del acompañamiento de CIDET a medida que avanza el proceso de acompañamiento.

En la Figura 10, se muestra la herramienta de diagnóstico desarrollado por CIDET para el proyecto E2 desarrollado en conjunto con INNPULSA. La herramienta ilustra la calificación de 180 preguntas que construyó CIDET para la valoración de las empresas. Las preguntas están enfocadas en las diferentes dimensiones anteriormente mencionadas. Así mismo las preguntas tienen 8 formatos diferente de repuesta que son convertidos a una escala general a través de un sistema de valoración común. La herramienta cuneta un sistema de filtrado por dimensiones, criterios y preguntas con el cual se busca facilitar a las empresas visualizar características particulares en las que quieran ahondar. También, se cuenta con formato de valoración poligonal con el que se muestra la ponderación total por dimensión para la empresa, lo que facilita para empresa y los evaluadores identificar cual es el aspecto en el cual la empresa es más débil y fuerte. Allí también está consignado el valor de LCOE - Levelized Cost of Energy, que en español significa el costo nivelado de la energía.

Figura 10. Herramienta de diagnóstico INNPULSA.

Principalmente para comparar directamente los costes de diferentes fuentes de energía. Al mirar el coste nivelado de cada una de dichas fuentes, tenemos un coste que está estandarizado. Es decir, podemos mirar

el LCOE de una instalación solar y compararlo directamente con el LCOE de una instalación térmica. El que sea menor de los dos nos estará diciendo cuál de las instalaciones genera energía más barata. En el DashBoard se presenta la suma de dos LCOE, el estado y usuario, el primero se refiere al costo que deberá asumir el estado (promotor gubernamental) para el proyecto y el segundo se refiere al costo que deberá asumir los usuarios (beneficiarios) del proyecto por recibir el servicio de energía, esta tarifa para el usuario es lo que se entiende como pago de factura de servicios.

También se presenta la TIR o tasa interna de retorno calculada para la empresa de acuerdo con su propuesta. Se presenta el número de beneficiarios para el proyecto y el valor en dinero de diferentes aspectos relacionados.

La herramienta ofrece además otras cualidades que ayudan tanto a los evaluadores como las empresas que deben diligenciar el cuestionario a hacer de una forma sencilla y rápida. En la pestaña instrucciones está cargado un documento en el cual está consignadas recomendaciones, ejemplos de llenado de la encuesta y finalidad de la herramienta. Por otra parte, se encuentra el cuestionario que es el se debe diligenciar, una pestaña de informa en la cual se presenta la calificación de cada pregunta, criterio y dimensión y finalmente una pestaña de CV (Convención de Valoración) en la que se visualiza los diferentes tipos de pregunta existente la forma y a la forma como se califican.

Código de Dimensión	Dimensión	Criterio de Evaluación	Código de Criterio	Preguntas	Años	Código de Pregunta	Escala
D1	Ambisión	Crecimiento	C1	**Crecimiento en ventas** Indique el incremento porcentual en ventas netas de los últimos 3 años (establezca la comparación con el año inmediatamente anterior).	2017	P1	0 A 10
					2018	P2	0 A 10
					2019	P3	0 A 10
				Crecimiento en Ganancias Indique el incremento porcentual en excedente neto de los últimos 3 años (establezca la comparación con el año inmediatamente anterior).	2017	P4	0 A 10
					2018	P5	0 A 10
					2019	P6	0 A 10
				Crecimiento en Clientes Indique el incremento porcentual en clientes nuevos de los últimos 3 años (establezca la comparación con el año inmediatamente anterior).	2017	P7	0 A 10
					2018	P8	0 A 10
					2019	P9	0 A 10
				Crecimiento en la participacion del mercado Indique el porcentaje de participación en el mercado en los últimos 3 años.	2017	P10	0 A 10
					2018	P11	0 A 10
					2019	P12	0 A 10
		Conformación y disposición de los equipos de trabajo	C2	Con respecto al personal interno de la empresa, indique el porcentaje de empleados que cuentan con formación académica y experiencia necesaria para desempeñar sus funciones.		P13	0 A 10
				La empresa motiva a los empleados a porponer soluciones y generar nuevas ideas.		P14	0 A 10
				El equipo de trabajo está dispuesto a correr riesgos con el fin de lograr los objetivos.		P15	0 A 10
		Disposición y asignación de recursos financieros destinados para el aumento de los niveles de crecimiento	C3	Indique el nivel de planificación del proceso de gestión financiera.		P16	0 A 10
				Crecimiento en inversiones equipos e infraestructura Indique el incremento porcentual de la inversión en crecimiento (adquisición de maquinaria, ampliación de planta, adquisición de filiales, entre otras) en los últimos 3 años, establezca la comparación con el año inmediatamente anterior.	2017	P17	0 A 10
					2018	P18	0 A 10
					2019	P19	0 A 10

Figura 11. Cuadro de codificación del DashBoard.

En la Figura 11, se presenta un fragmento del cuadro de codificación de preguntas, criterios y dimensiones. En este cuadro se encuentra el aspecto detallado para cada uno de los códigos mostrados en el DashBoard, de esta forma se interpreta

de forma más clara toda la codificación presentada y se puede identificar de formas muy precisa cual la calificación correspondiente a cada aspecto anteriormente mencionado.

Finalmente se resalta que, a través de este tipo de herramientas, las empresas pueden hacer uso de las técnicas de análisis de datos y conceptos de visualización de información dando valor agregado a los clientes y encaminándose hacia la digitalización y automatización de los procesos. De forma particular, CIDET ha logrado dar valor agregado al proceso de fortalecimiento de las empresas que participaron en la valoración, ofreciéndoles una forma interactiva de visualizar sus capacidades para el desarrollo del proyecto, lo que resulta en un proceso más ágil de retroalimentación.

3. Proyecto: campaña seguridad y riesgo eléctrico en Colombia.

CIDET como una de las entidades de certificación de productos eléctricos más importantes y reconocidas a nivel nacional en el sector eléctrico en Colombia ha identificado algunas falencias en la prevención y educación en relación con la seguridad y riesgo eléctrico. Debido a esto, la organización inició el proceso de planeación de un proyecto denominado "Reto: ¿Cómo aumentar la conciencia del ciudadano común sobre el riesgo de los sistemas eléctricos que consume?", este proyecto se va a ejecutar bajo la figura de campaña y tiene dos objetivos esenciales:

1) Concientizar a los ciudadanos sobre el riesgo de las instalaciones eléctricas domiciliarias, empresariales y públicas, cuando se tienen en cuenta las recomendaciones de manipulación o l calidad de los productos que instalan en las mismas.
2) Posicionamiento de CIDET como una de las principales entidades de certificación en Colombia.

Para el desarrollo del proyecto inicialmente se realizó un Benchmarking sobre el panorama de las campañas de riesgo eléctrico a nivel nacional e internacional. A través de esta evaluación se podrá identificar cuáles son las entidades que en Colombia y el mundo se han enfocado más en este tipo de proyectos, de esta forma se identificarán la forma como las ejecutan y el público al que desean llegar. Por otra parte, se desarrolló un DashBoard estadístico mediante el cual se logra visualizar el panorama de riesgo a nivel nacional.

Contexto Nacional

En Colombia las campañas de seguridad y riesgo eléctrico inicial en el 2014, como respuesta la aprobación en 2013 del Reglamento técnico de instalaciones eléctricas – RETIE, el que cual exige el certificado de conformidad para todos dispositivos empleados en instalaciones eléctricas. El certificado de conformidad (COC) es El certificado de producto es un documento mediante el cual una

tercera parte, independiente de una relación contractual cliente-proveedor, da constancia por escrito que un producto cumple con los requisitos establecidos en el RETIE. A partir de la aprobación del RETIE en Colombia se iniciaron una serie de actividades a nivel de leyes y campañas de concientización por parte de algunas empresas relacionadas con el sector eléctrico en Colombia. Este panorama se expone a continuación me diente la revisión de algunas leyes y campañas desarrolladas en el país:

Leyes

- Resolución Comisión de Regulación de energía y Gas (CREG) 123 de 2014. Por la cual se ordena la inclusión de información en las facturas del servicio, para promover el uso eficiente y el ahorro de energía eléctrica. Se invita a travésde un mensaje en la factura de servicios al manejo eficiente y responsable de la energía de modo que se contribuya a la sostenibilidad del medio ambiente.

- La ley de seguridad avanza en Córdoba a través de la resolución N°46/2017 (Empresa electro Instalador). Entra a regir la ley de seguridad eléctrica 10.281 a partir del 1 de diciembre de 2017. Ofrecen una cartilla amigable para orientar a las personas de cómo se debe hacer la certificación de su red eléctrica domiciliaría. También ofrecen noticias de accidentes con redes eléctricas de personas comunes. Fecha (15 de agosto de 2018, 27 de febrero de 2020).

- Conforme al artículo 78 de la Constitución Política. serán responsables de acuerdo con la ley, quienes en la producción y en la comercialización de bienes y servicios, atenten contra la salud, la seguridad y el adecuado aprovisionamiento a consumidores y usuarios.

Campañas

Las campañas relacionadas con la seguridad y riego eléctrico en Colombia son incentivadas principalmente por las empresas prestadoras de servicios públicos en Colombia, las entidades de seguros de vida y los hospitales. La ejecución de estas campañas se basa en la creación de material audio visual y de cartillas educativas que normalmente son publicadas en la página de la empresa prestadora de servicios o en YouTube. La mayoría de las campañas desarrolladas son promovidas por la Asociación Colombiana de Distribuidores de Energía – ASOCODIS. Una de las campañas más representativas creadas por esta entidad se llama "La historia del Verraco de los Verracos", campaña que consiste en una serie de videos a partir de los cuales se expone algunas de las malas prácticas en uso de las redes de energía eléctrica. A continuación, se muestran otras campañas desarrolladas por otras entidades en Colombia en torno al tema de la seguridad y el riesgo eléctrico.

- Ecopetrol. Divulgación a través de videos plataforma de YouTube. Tratan temas sobre los riesgos de la electricidad en el hogar y en otros entornos. Realizan una comparativa entre los riesgos en hogar vs los riesgos que tiene las personas en unas subestaciones, siempre haciendo énfasis en seguridad y el riesgo. Se apoya en imágenes de accidentes reales. (Fecha 9 mayo de 2014).

- Hospital San Vicente Fundación y EPM. Cartilla "Uso seguro de la energía eléctrica" publicado en su página web. Su temática se centra en el uso indebido de las redes de energía eléctrica. Se apoya en textos y animaciones. (Fecha diciembre 1 de 2019).

- CELSIA. Video publicado en su página web y en la plataforma de YouTube. Se da recomendaciones para el uso seguro de las redes de energía. El video se apoya en una animación. (Fecha 2 de enero de 2017).

- ESSA – Grupo EPM. Se publican campañas a través de informes en la página oficial (Conéctese a la buena energía). Estos documentos contienen recomendaciones, números donde se puede hacer denuncias, normativas, riesgos eléctricos a los que se está expuesto. Fecha (01 agosto de 2017, 6 octubre de 2014/2016/2018).

- Superintendencia de Industria y comercio / Protección del consumidor. Publica un segmento en su página dedicado a la protección del consumidor. Aquí se tratan temas de los productos en general, tales como; garantías, calidad, seguridad del producto, vehículos en otros. Se presentan una cartilla con recomendaciones relacionadas con calidad y garantía de cualquier clase de producto. También se realiza un video con el personaje "Tal cual".

Contexto Internacional

A nivel internacional las campañas sobre seguridad y riesgo eléctrico son desarrolladas por entidades como central de bomberos, empresas prestadoras de servicios, entidades de atención al consumidor y entidades especializadas en la seguridad y el riesgo eléctrico como es la Fundación Internacional de Seguridad Eléctrica (ESFI). La ESFI es el organismo encargado de gestionar y ejecutar las campañas relacionadas con la seguridad y el riesgo eléctrico en Estados Unidos. También, realiza una labor de vigilancia, esta actividad se refiera a toma de datos sobre los accidentes relacionados con redes eléctricas en el país analizarlos y utilizarlos para la divulgación y la toma de decisiones para intervenir sectores específicos a través de formación o campañas de seguridad. De forma general, países como Estados Unidos, México, Chile y Argentina prestan una mayor atención al aspecto de la seguridad y el riesgo eléctrico, esto se evidencia en el mayor número de campañas relacionadas con el este tema. Las campañas de seguridad en estos pises se basan en una interacción más cercana con las personas ya que las intervenciones (campañas) incluso se televisan, se hacen apartados en áreas de entretenimiento, se dedican páginas enteras a la divulgación y educación

sobre la seguridad y riesgo eléctrico. En Estados Unidos a través de la ESFI en las escuelas primaría se desarrollan actividades de educación a través de su página en la tiene una plataforma de juegos, cartillas para apoyar a los profesores y una aplicación en 3D para mostrar los riesgos eléctricos en el hogar (ESFI-3D, 2020). A continuación, se muestran algunas de las campañas relacionadas con la seguridad y el riesgo eléctrico:

Campañas

Canal Profeco (Procuraduría general del consumidor - México) / Revista del consumidor. Divulgación con videos a través de la plataforma de YouTube haciendo campaña sobre la importancia de la calidad de las instalaciones eléctricas en general y su riego para la salud. Para ello utiliza estudios, estadísticas, escenas, entrevistas en calle a las personas y fotos de accidentes y una inducción básica a lo que es un circuito eléctrico. "Seguridad en Productos: Instalaciones eléctricas [Revista del Consumidor TV 4.2]" y "Seguridad en Productos: Riesgos en la instalación eléctrica [Revista del Consumidor TV 40.1]". Fechas (8 de octubre de 2013, 31 de enero de 2013, 7 de diciembre de 2016).

La International Copper Association (ICA) en Latinoamérica. Lleva a cabo una campaña informativa y de seguridad que implica trabajar en conjunto con los organismos de cada país en lo referente a la seguridad de las instalaciones eléctricas.

Organismo Supervisor de la Inversión en Energía y Minería (Osinergimin) – Perú. Campaña a través de videos en la plataforma de YouTube "¿Cómo prevenir accidentes eléctricos en el hogar? "y "Prevención de accidentes eléctrico en el hogar – Animación". La temática se centra en métodos de prevención ante los riesgos de los sistemas eléctricos en el hogar y en otros lugares. Se apoya en animaciones he imágenes de situaciones en el hogar. (Fecha 30 de junio de 2016, 9 de mayo de 2019).

Edelaysen (Empresa de servicios eléctricos – Chile). Divulgación a través de medios de comunicación como internet, usando su página principal, en la cual publican un post con tips para la seguridad eléctrica. También, se comunica la campaña "asegura tu energía en verano" través de canal de televisión (noticias). Fecha (8 enero de 2020, 26 de enero de 2019).

Gobierno de Valdivia, Cooprel y Grupo SAES. La campaña consiste en la divulgación de recomendaciones hacia la comunidad respecto a riesgos eléctricos presentes tanto en el hogar como en el entorno de la cuidad. Además, se da un muy breve listado de los elementos que más fallan e las redes eléctricas domiciliarias (protecciones diferenciales, estructuras metálicas no aterrizadas y conductores no debidamente canalizados a través de tubería). La campaña "Uso seguro de energía en fiestas patrias" en forma de comunicado en una página de noticias regional (voceroregional.cl). Fecha (8 de septiembre de 2015).

Oklahoma Gas & Electric (OG&E). Campaña educativa para el uso adecuado de elementos eléctricos tanto dentro como fuera del hogar. Se apoya en un video animado. Fecha (30 junio de 2014).

State Farm Insurance Empresa de seguros – EE. UU. Se presenta en la página un informe con un listado de riesgos eléctricos en el hogar. Se menciona mas no se le hace énfasis a la recomendación de tener productos de calidad. Se hace un listado de elementos críticos para la seguridad en el hogar y se menciona las características más relevantes de estos para garantizar la seguridad. Fecha (2018).

Administración Nacional de Usinas y Transmisiones Eléctricas (UTE). Presenta un programa de inclusión social para los hogares que tienen deficiencias en sus instalaciones eléctricas. La campaña se denomina "Conéctate seguro" y ofrece toda una guía que in incluye los siguientes elementos; Importancia del proyecto, beneficios, tarifas, consultas, vecinos etc. Toda una orientación que permite educar de forma básica a las personas sobre la importancia de la seguridad eléctrica. Se apoya en material visual y lectura Fecha (19 de septiembre de 2019, 25 de marzo de 2019, 7 noviembre de 2018).

Grupo aseguradora La segunda - Argentina. Divulgación de videos educativos a través de la plataforma de YouTube haciendo campaña de cómo prevenir accidentes eléctricos en el hogar. Da recomendación del uso de elementos de protección como disyuntores y tapas para tomacorrientes. Video "Accidentes del Hogar: Electrocución". Fecha (17 de febrero de 2017).

Electro Sur este S.A.A – Perú. Presenta una campaña de divulgación a través de YouTube con la cual brinda educación a la familia sobre los diferentes riesgos eléctricos presentes en el hogar y como los pueden mitigar. Se apoyan en una animación. Fecha (12 de abril de 2018).

Asociación Peruana de Consumidores y Usuarios (ASPEC). Divulgación a través de medios televisivos (TV Perú) en una sección de alerta consumidor. Se recomendaciones de seguridad eléctrica en el hogar por parte de un experto en el tema. También, se mencionan la importancia de revisar la calidad de algunos elementos eléctricos. Fecha (18 de diciembre de 2013).

International Fire Protección Association (NFPA) – Estados Unidos. Cartillascon tips y listado de chequeo para los elementos eléctricos dentro del hogar. También se muestran una serie de videos, donde muestran buenas prácticas para el uso de elementos eléctricos. Fecha (2017/ 2018 /2019).

Algunas de las características generales de las campañas internacionales se muestran a continuación:

- Se resalta que algunas campañas no diferencian entre hogares y espacios más laboral diferentes a los espacios donde se realizan actividades relacionadas con la electricidad. Sin embargo, la mayoría de las campañas se enfocan en la seguridad dentro del hogar.

- Las campañas generalmente se publican en las páginas con informes extensos, divulgación gráfica no tan cuidada y son difíciles de encontrar ya que no están en medios donde la persona común se desplace o visitenormalmente.
- No se evidencia participación por parte de universidades en las campañas de seguridad eléctrica rastreadas.

- Las campañas normalmente son ejecutadas por entidades gubernamentales, empresas de servicios eléctricos y aseguradoras.

- México con la procuraduría federal del consumidor y el gobierno. Para resaltar: Se realiza un trabajo de retro alimentación preguntándole a las personas, ¿Que saben de la seguridad eléctrica?

Estadísticas

Como se mencionó anteriormente la ESFI se apoya en estadísticas para ejecutar las campañas de seguridad y riesgo eléctrico en Estados Unidos, esta misma estrategia fue implementada por CIDET para desarrollar el proyecto de seguridad y riesgo eléctrico para Colombia ya que actualmente en ninguno de los gestores de estas campañas ha publicado o evidenciado el uso de datos para la divulgación, desarrollo y ejecución de las campañas de seguridad.

En la Figura 12, se muestra la información recolectada para el análisis del panorama de relacionado con la seguridad y riesgo eléctrico en Colombia. La información fue tomada de la base de datos abiertos del gobierno de Colombia y hace parte del registro "Formato 9" diligenciado mensualmente por la Superintendencia de Servicio Públicos Domiciliarios. En el formato se inscriben todos los accidentes de origen eléctrico reportados por la empresas u hospitales en el país (GOV.CO, 2020). De los resultados se puede resaltar los siguientes aspectos:

- En el periodo 2010 a 2019 en Colombia han ocurrido 2833 accidentes de origen eléctricos, de los cuales 2466 accidentes fueron de personas de sexo masculino y 367 accidentes son personas de sexo femenino. Esto deja en evidencia que la población más vulnerable son los hombres, esto debe principalmente a que a que este género es quien desarrolla la mayoría de las actividades relacionadas con energía, refiriéndose a técnicos en el área o incluso en situaciones cotidianas como cambiar o reparar elementos en el hogar.

- Las edades en las que se presenta la mayor parte de accidentes eléctricos en Colombia son entre los 10 a 54 años y de forma particular las edades más afectadas son 1 a 9 años y entre los 28 y 38 años. También se destaca que la accidentalidad de las mujeres se da en los primeros años de vida (1 a 9 años) y en la etapa de productividad (18 en adelante) la accidentalidad se reduce significativamente.

- La accidentalidad para las diferentes edades ha mantenido la misma tendencia a lo largo del periodo 2010-2019, sin embargo, el índice de accidentes en las edades de 1 a 9 años ha mostrado una gran variabilidad en todos los años. Esto indica que los accidentes en edades tempranas se deben a eventos fortuitos o descuido de los padres y que por el contrario la tasa sostenida de muertes en edades entre 18 a 54 años se debe principalmente a desconocimiento, malas prácticas de las reglas de seguridad o exceso de confianza en el desarrollo de las actividades.

- También se detectó que el nivel de escolaridad es un factor que influye fuertemente en la cantidad de accidentes por contacto eléctrico. Se puede apreciar en la Figura 12 que la mayoría de los accidentes pertenecen a personas cuyo grado de escolaridad era muy bajo o incluso nulo (otro) y medida que el grado de escolaridad aumenta disminuye el número de accidentes. Este fenómeno se debe a dos factos: el primer factor está relacionado con la actividad desempeñada, esto se refiere a que las personas más capacitas son quienes menos contacto o menos actividades directas realiza con la red eléctrica; el segundo factor está relacionado con la conciencia del riesgo, esto implica que quien más capacitado está, más consciente del riesgo esta. Por lo anterior, se resalta la importancia de la educación y concientización.

- La mayoría de los accidentes relacionados con las redes de energía ocasionan quemaduras, tetanización, perdida miembros (otras) o la muerte. También ocasiona otro tipo de lesiones de menor impacto como: conjuntivitis o derivadas, electrolisis, fibrilación o traumatismos.

- La mayoría de los accidentes relacionados con contactos eléctrico se dan en actividades relacionadas con mantenimiento o en labores realizadas en redes de distribución de energía. La causa de los accidentes se debe principalmente a contactos indirectos y descargas atmosféricas, lo que indica que un gran porcentaje de accidentes se podría evitar si incorporan mejores prácticas en la construcción de las redes eléctricas, sus sistemas de puesta a tierra y protocolos de intervención (5 reglas de oro) de las redes.

- A nivel nacional, los departamentos con mayor número de accidentes relacionados con redes eléctricas son Valle del Cauca, Atlántico, Bolívar, Antioquia y Magdalena. Las zonas del caribe tienen un índice de accidentalidad considerablemente mayor debido a la mala calidad de redes eléctricas en estas zonas que se agrava con la tendencia a la ilegalidad debido a la gran cantidad de comunidades vulnerables 8de bajos recursos) que se establecen allí. En Antioquia, particularmente en Medellín, el gran número de accidentes se debe al contrabando en las redes de distribución.

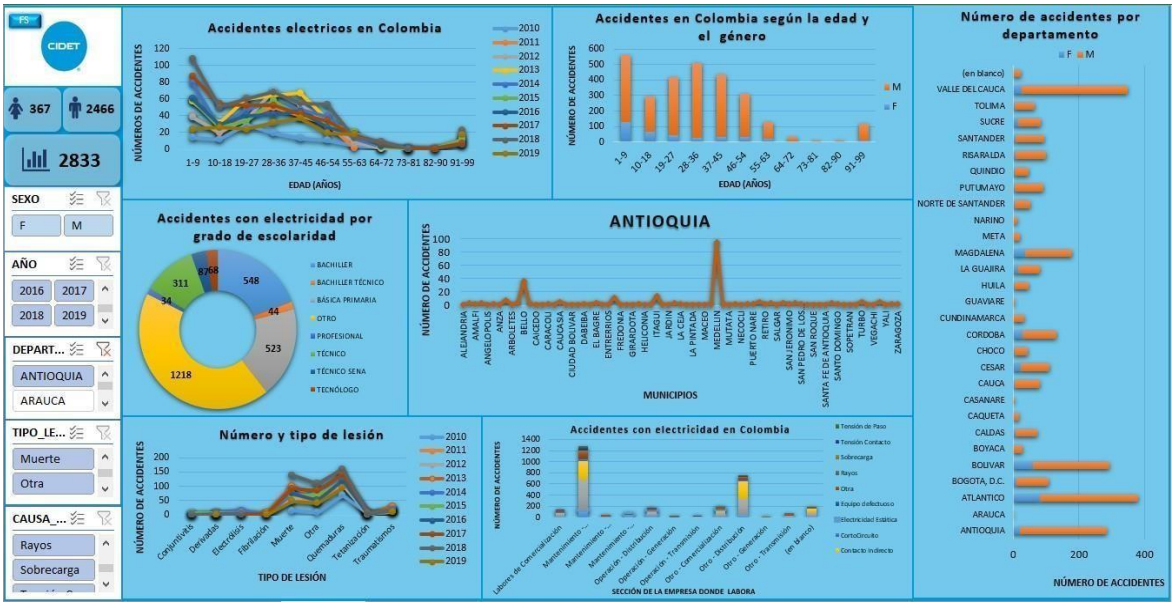


Figura 12. Estadísticas de accidentes eléctricos en Colombia 2010-2011.

Colombia debe trabajar más ampliamente en el desarrollo de campañas de concientización y educación del riesgo eléctrico ya que en relación con otros países se ha evidenciado un bajo interés en el tema. Para desarrollar estrategia de impacto social Colombia se puede apoyar en las experiencias de los grandes referentes internacionales en la ejecución de proyectos concientización de la seguridad y riesgo eléctrico. Uno de los referentes más importantes y en los cuales CIDET se apoyó para planear la campaña se describen brevemente a continuación.

Referentes

Estados Unidos y Europa. Estas zonas privilegian los estándares de calidad, eficiencia y seguridad de todos los productos importados y exportados. Para garantizar la seguridad ante el riesgo eléctrico de productos o instalaciones se imponen fuertes políticas de calidad a través de normativas internacionales como; IEEE, IEC. Para estados Unidos aplican normas como; ANSI, UL, NEMA, CPSC y ASM. Para la zona europea aplican normas como; DIN, BS, Cenelec y CEN.

Estados Unidos a través de NEMA, UL y CPSC creo La Fundación Internacional de Seguridad Eléctrica (ESFI). Es una organización sin fines de lucro fundada en 1994 que se dedicada exclusivamente a promover la seguridad eléctrica en el hogar y en el lugar de trabajo. Dentro de su página se puede identificar estadísticas, campañas de prevención del riesgo eléctrico, discusiones sobre el tema de la seguridad, recomendaciones etc. Se registran más de 9 campañas desde el 2013 (ESFI, 2020).

En la Figura 13, se muestra el número de campañas relacionados con la seguridad y riesgo eléctrico mapeadas a nivel nacional (azul) e intencional (naranjado). Se puede apreciar que el número de campañas a nivel internacional va en aumento

en los últimos años. A nivel nacional las campañas de seguridad y riesgo eléctrico iniciaron en el año 2014, esto se debe a que en el año 2013 en Colombia se aprobó el RETIE, lo que obligó a las empresas prestadoras de servicios públicos exigir certificados de calidad para las instalaciones eléctricas y los componentes de esta.

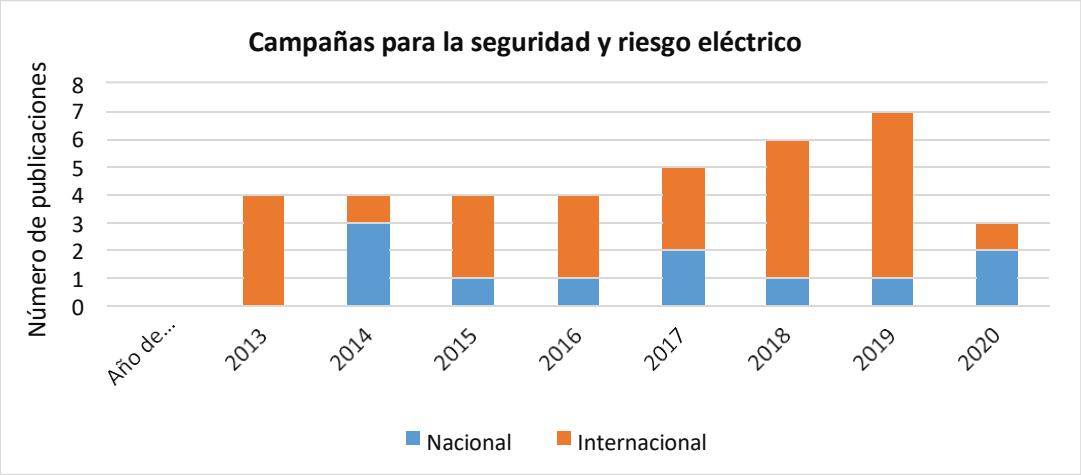


Figura 13. Numero de campañas sobre el riesgo eléctrico en Colombia en los últimos 8 años.

4. Proyecto: desarrollo de un estudio de mercado mediante el uso de técnicas de analítica de datos e inteligencia competitiva, con el fin de identificar oportunidades en el mercado.

Uno de los servicios que CIDET ha incorporado más recientemente es el desarrollo de estudios de mercados. Esta nueva incursión se debe a que este tipo de estudios tiene una función esencial en el mundo de las ventas, son los que permiten a las empresas determinar si su idea de negocio es viable o no. A través de este proceso se logra determinar los puntos clave que le van a permitir el éxito de un proyecto empresarial. Los estudios de mercado también tienen como objetivo permitirles a las empresas identificar las oportunidades relacionadas con la innovación en productos y servicios, lo que se traduce para la empresa y el país como progreso. El gran impacto que significa el desarrollo de un estudio de mercado para una empresa fue lo que motivo CIDET a iniciar con la ejecución de este tipo de proyectos ya que uno de sus objetivos como organización es impulsar el desarrollo empresarial y em particular el del sector eléctrico.

Un estudio de mercado puede desarrollarse de múltiples maneras, ya que tiene diversas formas de segmentación. Sin embargo existen dos clasificaciones generales: la primera se denomina estudio de mercado primario, en este tipo de estudios se desarrollan actividades tradicionales pero confiables como los focus group (grupos de concentración o prueba), encuestas, entrevistas, investigaciones

de campo y observaciones del producto o punto de venta; la segunda se denomina estudio de mercado secundario, en este caso el estudio se desarrolla a partir de datos secundarios obtenidos de otras fuentes que son aplicables al producto de interés. La razón para acceder a un estudio de mercado secundario se debe a que son relativamente menos costos que el estudio mercado primario. Por otra parte, este tipo de estudios tiene la desventaja de que los resultados no son específicos al área de investigación, debido a que los datos utilizados vienen de tendencias y son difíciles de validar. El estudio de mercado desarrollado por CIDET tiene un componente primaría y una componente segundaría. La componente primaría del estudio de mercado se basa en el uso bases de datos relacionadas fuertemente con el sector eléctrico en particular, mientras que la componente secundaria del estudio se refiere a un conjunto de suposiciones que se usan para determinar algunas cantidades o tendencias para el estudio de mercado.

A través del estudio de mercado desarrollado por CIDET se quiere responder a preguntas base como: ¿Qué sucede en el mercado?, ¿Cuáles son las tendencias?, quienes son los competidores de la marca o servicio? La respuesta a estas preguntas tiene como objetivo identificar las necesidades y objetivos de los diferentes segmentos de mercado, determinar su división, determinar costos asociados, identificar las limitaciones y desventajas de los productos y por último determinar los agentes que influyen su dinámica.

Una de las disciplinas más usadas en el campo del marketing, ventas y estudios de mercado es la ciencia de datos, esto se debe a que a través de estas técnicas se pueden realizar predicciones con cierto nivel de certeza, lo que es muy provechoso y ventajoso a la hora de tomar decisiones acertadas. En el desarrollo de este tipo de estudios CIDET ha adoptado algunas de estas técnicas para realizar proyecciones del mercado. Una de las técnicas usadas por la empresa es la regresión lineal. Esto no significa que solo use este tipo de regresiones, ya que el grado de la regresión está estrechamente relacionado con el comportamiento de los datos de interés, esto quiere decir que previo a seleccionar la técnica se observa la dinámica de los datos y posteriormente se selecciona la técnica más apropiada. En este caso en particular se llevará a cabo un ejemplo a través del cual se mostrará todo el proceso de análisis al cual se someten los datos en la construcción de un estudio de mercado empleando un modelo lineal.

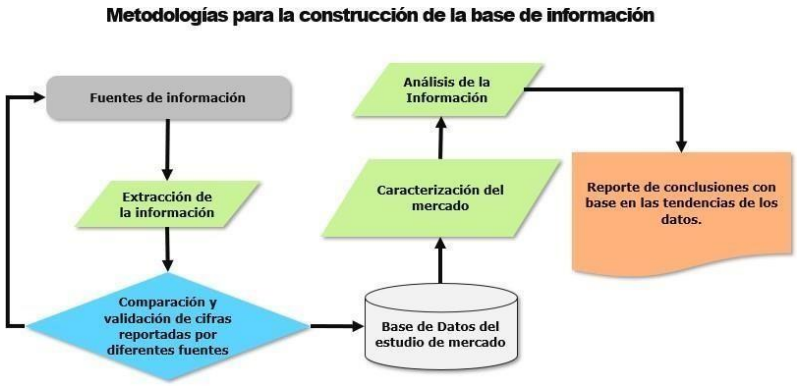


Figura 14. Metodología para la construcción de una base de datos.

Una de las primeras tareas que se deben llevar a cabo en todo proceso de construcción de un modelo de predicción es la recolección y limpieza de datos. En este proceso se busca que los datos que se espera usar para la construcción del modelo estén en un formato común, no existan valores atípicos, valores no válidos, valores en blanco etc. Como se muestra en la Figura 14, proceso de construcción de la base una base datos inicial con la identificación de todas las fuentes de información necesarias para desarrollar el modelo, por lo general en la construcción de modelos de proyección se suele tener una variable de interés otras relacionadas. Para la construcción de modelos lineales generalmente no es necesario una fuente de datos muy extensa, sin embargo, si se debe realizar un trabajo riguroso en la selección de los pocos datos que se incorporen a la base de datos de forma que el modelo pueda ser lo más representativo posible. En la extracción de los datos se busca llevar a una sola plataforma o software los valores, de forma que se pueda realizar una limpieza general de los mismo y lo que resulta en una manipulación más sencilla de los mismo. Cuando existen varías fuentes de información es recomendable realizar un cruce de valores para detectar posibles inconsistencias en la data recolectada, si se ha verificado la calidad de la información entonces se procede a guardar esta como base para el modelo, de lo contrario se recomienda ponerla a parte (sin desecharla) ya que servirá para filtrar información nueva. Cuando se ha consolidado toda la información necesaria en la base de datos entonces se procede realizar una caracterización de esta de acuerdo con las características del estudio de mercado. En esta etapa se etiqueta la información de nuestra base de datos, es decir, construirnos la tabla de variables relacionadas con nuestra variable de interés y les asignamos los nombres adecuados. Es paso siguiente es la construcción del modelo para realizar el análisis respectivo de acuerdo con las necesidades y objetivos particulares. Por último, con base a lo observado en el punto anterior se realiza un banco de recomendaciones y conclusiones para el estudio de mercado.

Con fin de ilustrar el proceso anteriormente mencionado a continuación se realizará todo el proceso de construcción de un estudio de mercado muy básico a través de un ejemplo general. Además, con el ejemplo se quiere mostrar el uso

de algunas herramientas muy útiles para el desarrollo de este tipo de estudios como lo son Python y Excel.

Ejemplo

El modelo predicción de regresión lineal permite hallar el valor esperado de una variable aleatoria Y cuando X toma un valor específico, de modo que, las variaciones en X puedan explicar los cambios en Y. La aplicación de este método implica un supuesto de linealidad cuando la variable de interés presenta un comportamiento creciente o decreciente, por tal razón, se hace indispensable que previo a la selección de este método exista un análisis de regresión que determine la intensidad de las relaciones entre las variables que componen el modelo. Para explicar mejor esta definición se usa la siguiente situación enfocada en un ambiente empresarial.

Una empresa de consolas de juegos, la cual ofrece varios tipos de consolas en el mercado local requiere hacer algún tipo de proyección sobre el número de unidades mensuales que la empresa podrá vender de una consola en particular en particular en el próximo semestre.

Con el fin de cumplir con este objetivo, es preciso construir un modelo de regresión lineal. Para desarrollar este proceso es fundamental llevar a cabo los siguientes cuatro pasos.

1. Modelamiento: Aquí se desarrolla un modelo de regresión que consiste en generar una ecuación que sea el equivalente de nuestras variables de interés.

Para la compañía de consolas modelar o desarrollar un modelo de regresión, debe representar su conocimiento y creencias sobre el proceso en las ecuaciones. Basándose en la experiencia, se sabe que las ventas mensuales de unidades dependen de tres variables importantes. El precio al que se vende la consola, la cantidad mensual que la empresa gasta en publicidad de la consola y la cantidad mensual que se gasta en promociones de esta. Sin embargo, estas no son las únicas variables que influyen en la venta de los juguetes. También se podría incluir variables como: intensidad de la competencia, interés de los consumidores por la consola, época del año. La cantidad de variables puede ser extensa incluso puede que jamás podemos incluir o pensar en todas.

Para este caso se decide usar tres variables como variables explicativas para modelar las ventas mensuales de una consola.

$$\boldsymbol{Ventas} = \boldsymbol{\beta_0} + \boldsymbol{\beta_1} * \boldsymbol{Precio} + \boldsymbol{\beta_2} * \boldsymbol{Gpublicidad} + \boldsymbol{\beta_3} * \boldsymbol{Gpromociones} \quad (1)$$

Donde:

Ventas: Unidades vendidas por mes de una consola.

Esta variable es nuestra variable de interés, será la que le va a dar respuesta a nuestro problema. Esta variable recibe normalmente varios nombres: la variable de

interés, la variable Y, la variable dependiente, la variable de respuesta, la variable regresiva, la variable del lado izquierdo.

Del lado derecho de la ecuación (1) están las variables que explican el cambio en las ventas de las consolas.

- **Precio:** Precio de la consola en un mes determinado.
- **Gpublicidad:** Gasto mensual en publicidad.
- **Gpromociones:** Gasto mensual en promociones.

Estas variables son las que influyen en el comportamiento de nuestra variable de interés. También son conocidas por varios nombres: las variables X, las variables independientes, las variables covariantes, las variables regresivas.

$\boldsymbol{\beta_0, \beta_1, \beta_2, \beta_3}$: Son constantes o coeficientes del modelo de regresión.

Estos coeficientes o parámetros del modelo de regresión nos indican el impacto que tiene cada una de las variables explicativas (x) del modelo en la variable de interés (y). Estos valores deben ser estimados (en este caso lo haremos usando Python y Excel).

Por último, se aclara que el modelo de regresión siempre se escribirá el en el formato de la ecuación (1).

2. Estimación: Para estimar el modelo creado entonces se usa algún software, en este caso se usará Python y Excel para tal fin.

Las regresiones lineales se dividen en dos categorías:

- Regresión lineal Simple: Cuando existe una única variable explicativa (x) para el modelo.

- Regresión lineal Múltiple: Cuando existen dos o más variables explicativas (x) para el modelo de interés.

Es importante saber que se necesita conocer datos de todas las variables explicativas que intervienen en el modelo lineal creado (ecuación 1) para generar una estimación de este.

Metodología usando Excel

Ahora el primer paso es dirigirnos a nuestra lista de datos de Excel, como se muestra en la Figura 15.

	A	B	C	D	E
1	Month	Unit Sales	Price ($)	Adexp ('000$)	Promexp ('000$)
2	1	73959	8,75	50,04	61,13
3	2	71544	8,99	50,74	60,19
4	3	78587	7,50	50,14	59,16
5	4	80364	7,25	50,27	60,38
6	5	78771	7,40	51,25	59,71
7	6	71986	8,50	50,65	59,88
8	7	74885	8,40	50,87	60,14
9	8	73345	7,90	50,15	60,08
10	9	76659	7,25	48,24	59,90
11	10	71880	8,70	50,19	59,68
12	11	73598	8,40	51,11	59,83
13	12	74893	8,10	51,49	59,77
14	13	69003	8,40	50,10	59,29
15	14	78542	7,40	49,24	60,40
16	15	72543	8,00	50,04	59,89
17	16	74247	8,30	49,46	60,06
18	17	76253	8,10	51,62	60,51
19	18	72582	8,20	49,78	58,93
20	19	69022	8,99	48,60	60,09
21	20	76200	7,99	49,00	61,00
22	21	69701	8,50	48,00	59,00
23	22	77005	7,90	54,00	59,50
24	23	70987	7,99	48,70	58,00
25	24	75643	8,25	50,00	60,50
26					

Sheet1 Sheet2 Sheet3

LISTO

Figura 15. Lista de datos para las variables descriptivas del modelo.

Luego nos dirigimos a la pestaña de datos y hacemos Clic en la opción "Análisis de datos", como se muestra en la Figura 16.

Figura 16. Métodos de análisis de datos de Excel.

Luego seleccionamos el método de regresión, como se muestra en la Figura 17.

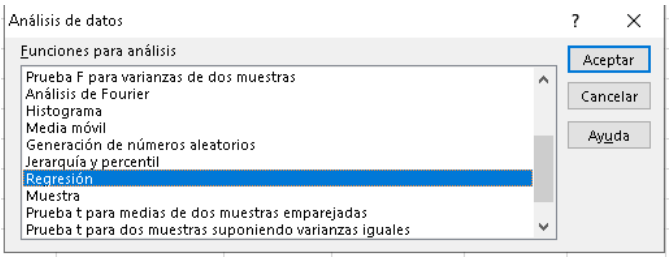


Figura 17. Método de regresión Excel.

Ahora se nos despliega el cuadro de la Figura 18, aquí seleccionaremos inicialmente la opción Rótulos ya que nuestro registro de datos tiene encabezados. Para el input de datos "Rango Y de entrada" seleccionaremos todos los datos incluyendo el encabezado de la columna "Unit Sales" que se puede observar en la Figura 15 Ya que esta es nuestra variable de interés (Y).

Luego en el campo "Rango X de entrada" ingresamos las columnas de datos relacionadas con el precio (precio), gasto publicidad (Adexp) y gasto en promociones (Promexp) que son nuestras variables descriptivas (X). Aquí es importante mencionar que las columnas de datos de las variables descriptivas

deben estar una al lado de la otra para que el análisis hecho por Excel se ejecute correctamente.

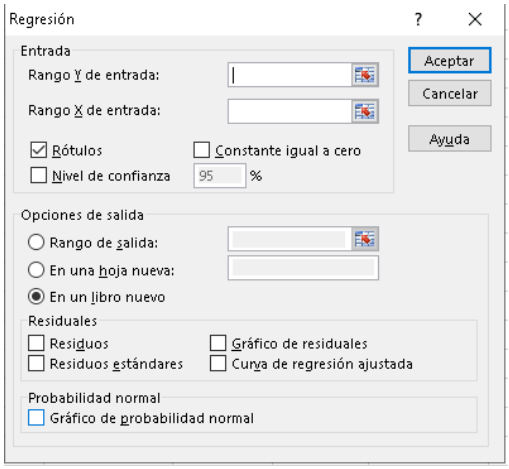


Figura 18. Parametrización de la regresión.

Le damos aceptar y Excel nos genera un informe con todas las características de la regresión, como se muestra en la Figura 19.

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,9267387
Coeficiente de determinación R^2	0,8588447
R^2 ajustado	0,8376714
Error típico	1274,9355
Observaciones	24

ANÁLISIS DE VARIANZA

	Grados de liberta	*ma de cuadrados*	*medio de los cuadrad*	*F*	*Valor crítico de F*
Regresión	3	197798832,8	65932944,28	40,5626221	1,08482E-08
Residuos	20	32509212,11	1625460,605		
Total	23	230308045			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95,0%*	*Superior 95,0%*
Intercepción	-25096,83	24859,61131	-1,009542451	0,324773156	-76953,07343	26759,40758	-76953,07343	26759,40758
Price ($)	-5055,27	526,3995537	-9,603484331	6,21954E-09	-6153,320094	-3957,219638	-6153,320094	-3957,219638
Adexp ('000$)	648,61214	209,0048787	3,103334928	0,005602344	212,635603	1084,588678	212,635603	1084,588678
Promexp ('000$)	1802,611	392,8485427	4,588564702	0,000178016	983,1432557	2622,078657	983,1432557	2622,078657

Figura 19. Resumen de análisis de regresión de Excel.

En la Figura 19, vemos que el análisis de Excel nos ofrece los valores para los coeficientes o constantes del modelo de regresión a partir de los cuales podremos medir el impacto de las variables descriptivas (X) en la variable de interés (Y). Los valores de los coeficientes son los siguientes:

$$\boldsymbol{Intercepción} = \boldsymbol{\beta}_0 = -25096{,}83$$

$$\boldsymbol{Price}\ (\$) = \boldsymbol{\beta}_1 = -5055{,}27$$

$$\boldsymbol{Adexp}\ (\boldsymbol{'000\$}) = \boldsymbol{\beta}_2 = 648{,}6121$$

$$Promexp\ ('000\$) = \beta_3 = 1802{,}611$$

Estas constantes corresponden respectivamente a constate independiente de la ecuación (B0), Impacto del precio mensual de la consola (B1), impacto del gasto en publicidad (B2) y el impacto de gasto en promociones (B3). Por lo que nuestra ecuación para el modelo queda de la siguiente forma.

$$Ventas = \beta_0 + \beta_1 * Precio + \beta_2 * Gpublicidad + \beta_3 * Gpromociones \quad (1)$$

$$Ventas = -25096{,}83 - 5055{,}27 * Precio + 648{,}6121 * Gpublicidad + 1802{,}611 * Gpromociones \quad (2)$$

Metodología usando Python

```
# -*- coding: utf-8 -*-
"""
Created on Sun May 17 08:05:12 2020
@author: León B
"""
# -------------------------- Librerías ----------------------
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.linear_model import LinearRegression
from sklearn.metrics import mean_squared_error
# ---------------------- Estilo de gráfica ----------------------
# Para dar estilo personalizado a la grafica usamos Seaborn
sns.set_style('darkgrid')
# ---------------------- Lectura de datos ----------------------
data = pd.ExcelFile("Base de Datos Empresa de consolas.xlsx")
print(data.sheet_names)
df = data.parse("Sheet1")
print(df.head())
# ---------------------- Lista de datos ------------------------
# Si se requiere contar los datos u organizar de mayor - menor o menor - mayor ( frecuencia )
Ounit_v = df["Unit Sales"].value_counts().sort_index(ascending = False)
# Variable de interes - Unidades vendidas
unit_v = df["Unit Sales"].values
# Variables descriptivas - Meses, Precio, Gasto en Publicidad, Gasto en Promociones
month = df["Month"].values
price = df["Price ($)"].values
g_publ = df["Adexp ('000$)"].values
g_prom = df["Promexp ('000$)"].values


# ------------------ Arreglos de la regresión # 1 ----------------------
# ---------------------------------------------------------------------
#Variables Independientes o descriptivas
X = np.array([price, g_publ, g_prom]).T
Y = np.array(unit_v)
# ------------------- Modelo del problema ----------------------
regr = LinearRegression()
regr = regr.fit(X, Y)
# Para generar una predicción del modelo
Y_p = regr.predict(X)
# Valor de los coeficientes
Coef   = regr.coef_
B1 =  regr.coef_[0]
B2 =  regr.coef_[1]
B3 = regr.coef_[2]
# Valor  de intercepto
```

```
Inter = regr.intercept_
# Para calcular el error
error = np.sqrt(mean_squared_error(Y, Y_p))
R2 = regr.score(X, Y)
# ---------- Imprimir resultado del modelo # 1 ----------
print (" ---- Modelo # 1 ----- ")
# Parametros
print ('Error :', error )
print ('Coficiente R^2 :', R2)
print ('Coficiente :', Coef)
print ('Intercepto :', Inter)
# Ecuación del modelo
print ('Ecuación General de modelo: \n Y = {0} {1} * Price + {2} * Gpublicidad + {3} * Gpromociones'.format(Inter, B1, B2, B3))
# ---------- Evaluación del modelo # 1 ----------
print (" ---- Escenarios / modelo # 1 ----- ")
# Se evaluan 3 escenarios
price_s = [9.10,7.10,8.1]
g_publ_s = [52,48,50]
g_prom_s = [61,57,60]
for i in range(0, len(price_s)):
  unit_v_s = Inter + B1 * price_s[i] + B2 * g_publ_s[i] + B3 * g_prom_s[i]
  print ("Escenario {0}".format(i))
  print ("Unit Sales: ", unit_v_s)
# ---------- Gráficas 1 ----------
fig_1 = plt.figure()
plt.plot(price, unit_v, "o")
plt.xlabel("Price")
plt.ylabel("Unit Sales")
plt.title("Unit Sales vs Price")
fig_1.show()
# ---------- Graficas 2 ----------
fig_2 = plt.figure()
plt.plot(g_prom, unit_v,      'ro')
plt.xlabel("Promotional Investment")
plt.ylabel("Unit Sales")
plt.title("Unit Sales vs G prom")
fig_2.show()
# ---------- Graficas 3 ----------
fig_3 = plt.figure()
plt.plot(g_publ, unit_v,      'go')
plt.xlabel("Marketing Investment")
plt.ylabel("Unit Sales")
plt.title("Unit Sales vs G publ")
fig_3.show()
```

Para el ejercicio en Python se usa la librería Sklearn que es una de las librarías de mayor uso para Machine Learning e Inteligencia artificial. De esta libraría usa el método LinearRegression para calcular los coeficientes del modelo de regresión lineal. Luego de ejecutar el código presentado anteriormente se puede apreciar la ecuación con sus respectivos coeficientes B0, B1, B2 y B3, como se muestra en la Figura 20. Además, se puede apreciar el coeficiente de determinación R^2 y el error típico del modelo. Esto confirma su consistencia con el modelo obtenido mediante Excel.

```
 ---- Modelo # 1 -----
Error : 1163.8515818351275
Coficiente R^2 : 0.8588446525398427
Coficiente : [-5055.26986592   648.61214026  1802.61095612]
Intercepto : -25096.832921870096
Ecuación General de modelo:
 Y = -25096.832921870096 -5055.2698659208445*Price + 648.6121402597208*Gpublicidad + 1802.6109561246005*Gpromociones
```

Figura 20. Resumen de análisis de regresión de Python.

De este modo se completa la estimación del modelo y se puede seguir al siguiente paso que es la inferencia.

3. Inferencia: En esta etapa se le da una interpretación a la estimación del modelo de regresión.

Antes de empezar se hace énfasis en que un modelo se puede escribir de forma general como:

$$\boldsymbol{Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_k * X_K} \quad (3)$$

Donde:

$\boldsymbol{\beta_0}$= Intercepto

$\boldsymbol{\beta_1, \beta_2, \beta_3, \beta_k}$ = Coeficientes que indican el impacto de las variables independientes sobre la variable de interés y

$\boldsymbol{X_1, X_2, X_3, X_K}$ = Son las variables que describen el a la variable de interés y

$\boldsymbol{Y}$ = Es la variable de interés

Para la interpretación de estas las variables del modelo de la ecuación (3) se usará como ejemplo B1.

Cuando la variable X1 aumenta en una unidad entonces Y incrementará en B1 unidades y todas las demás variables del modelo se mantienen al mismo nivel.

1. Debe tenerse muy en cuenta que X1 aumenta en una unidad, no un porcentaje, así que, si la unidad de X1 es un kilogramo, entonces un incremento de una unidad implica que X1 se incrementa en un kilogramo. Si la unidad es el 1.000 kilogramo, entonces un incremento de una unidad implica que X1 se incrementa en 1.000 kilogramos.

2. Por otra parte, la interpretación dice que Y aumenta en una unidad B1 y esta tiene las mismas las unidades de la variable Y. Así que, por ejemplo, si la variable Y se mide en términos de millones de dólares, entonces lasunidades B1 implica B en millones de dólares.

3. Por último, es importante aclarar la última parte de la interpretación, que dice que todas las variables se mantienen al mismo nivel. Esto es importante cuando se quiere interpretar el impacto de una determinada variable exponencial en la variable Y.

Teniendo en cuenta la información anterior, a continuación, se interpretará los coeficientes calculados para el modelo de ventas de consolas.

$$\boldsymbol{Ventas} = -25096{,}83 - 5055{,}27 * \boldsymbol{Precio} + 648{,}6121 * \boldsymbol{Gpublicidad} + 1802{,}611 * \boldsymbol{Gpromociones} \quad (2)$$

- B1: el valor estimado de este coeficiente es un valor negativo de 5055.27. La interpretación genérica del coeficiente es que por cada unidad de aumento en la variable X, la variable Y aumenta en unidades B. Todas las demás variables permanecen en el mismo nivel. Así que, en este caso particular, la interpretación se traduce en que cuando el precio de la consola aumenta en una unidad,

que es un dólar, entonces las ventas, que es mi variable Y, disminuye en 5055,27 unidades.

- B2: este coeficiente mide el impacto de los gastos de publicidad. Entonces, por cada unidad de incremento en el gasto publicitario, en este caso, la unidad de gasto publicitario es de 1.000 dólares, como se muestra en la tabla de datos de la Figura 15 (Adexp). La interpretación es la siguiente: por cada $ 1,000 de incremento en el gasto publicitario, las ventas unitarias por mes se incrementan en 648.61 unidades. Todas las demás variables permanecen al mismo nivel. Implicando que, si no cambiamos el precio, no cambiamos los gastos de promoción, los mantenemos al mismo nivel que están, y sólo aumentamos los gastos de publicidad en $1,000, entonces esperamos que las ventas unitarias aumenten en 648, 0.61 unidades. Por supuesto, no se puede tener 0,61 unidades, por lo que se redondea a un número menor y se diría que se espera que las ventas unitarias aumenten en 648 unidades. Además, esto también significa que si los gastos de publicidad se incrementan en 10.000 dólares. Entonces esperaríamos que el ahorro de unidades aumente en 6.486,12 unidades. Todas las demás variables permanecen al mismo nivel. Dado que esta es una ecuación lineal, cada interpretación puede construirse de forma independiente ya que la suma de sistemas lineales es un sistema lineal.

- B3: el coeficiente mide el impacto de los gastos en promociones. También puede ser interpretado de manera similar. El valor del coeficiente es 1802,65 positivo. Los gastos de promoción están en unidades de medida de 1.000 dólares. Esto significa que, por cada $ 1,000 de incremento en los gastos de promoción, esperaríamos que las ventas unitarias aumenten en 1802.61 unidades. Todas las demás variables permanecen al mismo nivel. Es decir, si el precio y el gasto de publicidad se mantienen al mismo nivel, no cambiaran y sólo aumentarán los gastos de promoción en 1.000 dólares, se esperaría que las ventas por unidad aumenten en 1802,61 unidades. Nuevamente, esto significa que si los gastos de promoción aumentan en $100, esperarías que las ventas por unidad aumenten en 180,26 unidades, manteniendo las otras variables en el mismo valor.

- B0: es el valor de la variable Y cuando todas las variables X son cero. En otras palabras, este será el valor de nuestra variable de interés cuando todas las variables explicativas son cero. Esta es la interpretación técnica de la beta cero. Sin embargo, esta interpretación técnica puede o no tener una interpretación relevante desde el punto de vista administrativo de una empresa. Sin embargo, para este ejercicio, implica que el valor de las ventas unitarias sería un valor negativo 25096.83, cuando todas mis variables X son cero. Cuando el precio es cero, el gasto en publicidad es cero, y el gasto en promoción también es cero.

Esto es una interpretación técnica de B0. En este caso, esta interpretación técnica no tiene una relevancia gerencial, ¿por qué?:

Se hablando de una situación en la que se está vendiendo una consola gratis y luego se calcula cuáles serían las ventas por unidad, no tiene sentido desde el punto de vista gerencial ya que se podría pensar este número negativo indica las unidades que estoy dejando de vender, pero esto es una interpretación errónea. Por tal motivo, se recomienda tener mucho cuidado en la interpretación de B0 de forma gerencial. Debe preguntarse si tiene sentido gerencial hablar de la situación en la que todas mis variables X son cero. En este caso en particular, claramente no tiene sentido gerencial hablar de una situación en la que estoy regalando mi consola. Estoy incurriendo en cero gastos en publicidad y gastos en promociones. Para este tipo de situación se dice que el coeficiente B0, en este casi tiene una interpretación técnica que se requiere para ajustar el modelo a los datos.

4. **Predicción:** En esta última etapa se construye una predicción para nuestra variable de interés. Para ellos se construyen tres escenarios hipotéticos que buscarán predecir el número de ventas de unidades de consolas para los próximos 6 meses.

Escenario 1: el primer escenario tiene un precio alto de 9.10 dólares. Pero respaldado con mayores niveles de publicidad y gastos de promoción. En particular, un presupuesto mensual de gastos de publicidad de 52.000 dólares, y un presupuesto mensual de gastos de promoción de 61.000 dólares.

Escenario 2: un bajo precio de 7,10 dólares para penetrar en el mercado, pero también tiene bajos niveles de publicidad mensual, y gastos de promoción. En particular, un nivel mensual de 48.000 dólares para la publicidad, y 57.000 dólares para las promociones.

Escenario 3: este escenario es más bien un escenario intermedio, donde la compañía pondrá el precio de la consola a 8.10 dólares, y asignará una cantidad mensual de 50.000 dólares por publicidad y una cantidad mensual de 60.000 dólares por promociones.

Suponiendo que el objetivo de la compañía es maximizar la venta de unidades de juguetes sin importar nada más. Con base en este objetivo se hará el cálculo para cada una de estas situaciones de forma que nos permita dar una recomendación. Para ello usaremos Excel y Python para evaluar la ecuación del modelo para cada uno de los escenarios.

$$\boldsymbol{Ventas} = -25096{,}83 - 5055{,}27 * \boldsymbol{Precio} + 648{,}6121 * \boldsymbol{Gpublicidad} + 1802{,}611 * \boldsymbol{Gpromociones} \quad (2)$$

Excel.

En Excel se construye para cada uno de los escenarios una ecuación por celdas usando el método de inserción de instrucciones "=", como se muestra en la figura 7. Es importante recordar que las unidades de gasto en publicidad (Adexp) y el gasto en promociones (Promexp) están en 1.000 $. Tomando como ejemplo el escenario uno, la forma correcta ingresar los valores sería la siguiente: Adexp = 52 y Promexp = 61. Recuerde que para dejar un valor fijo de una celda en Excel solo debe poner un símbolo de "$" luego de la letra que identifica el rango de cada celda. De esta forma podrá copiar la ecuación en las otras celdas sin volver a construirla. Este proceso se muestra en la Figura 21. En la Figura 22, se muestra la predicción del modelo para cada uno de los escenarios propuestos.

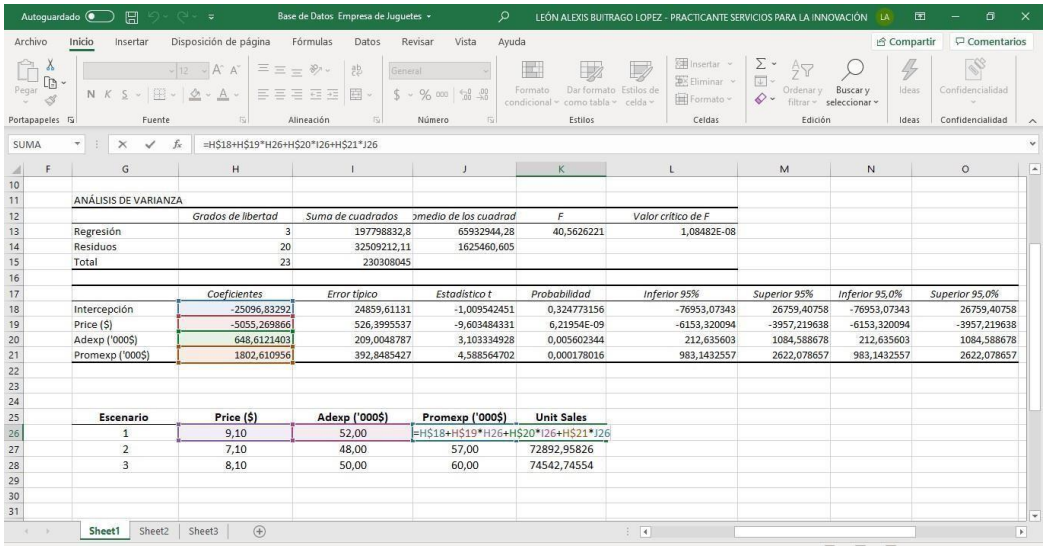


Figura 21. Ecuación para calcular una predicción con el modelo lineal calculado usando Excel.

Escenario	Price ($)	Adexp ('000$)	Promexp ('000$)	Unit Sales
1	9,10	52,00	61,00	72587,31092
2	7,10	48,00	57,00	72892,95826
3	8,10	50,00	60,00	74542,74554

Figura 22. Predicción de la venta de unidades de juguetes en tres escenarios diferentes evaluando el modelo en Excel.

Python

```
# − − − − − − − − − − − − − − − − − − Evaluación del modelo # 1 − − − − − − − − − − − − − − − − − −
print (" − − − − Escenarios / modelo # 1 − − − − − ")
# Se evaluan 3 escenarios
price_s = [9.10,7.10,8.1]
g_publ_s = [52,48,50]
g_prom_s = [61,57,60]
for i in range(0, len(price_s)):
  unit_v_s = Inter + B1 * price_s[i] + B2 * g_publ_s[i] + B3 * g_prom_s[i]
  print ("Escenario {}". format(i))
  print ("Unit Sales: ", unit_v_s)
```

Como se puede apreciar en la Figura 23 el modelo desarrollado en Python (se debe agrega este código al segmento de código presentado anteriormente) da como resultado las mismas unidades de consolas vendidas calculadas a través de Excel para los tres escenarios hipotéticos descritos anteriormente.

```
 ---- Escenarios / modelo # 1 -----
Escenario 0
Unit Sales:  72587.31091535633
Escenario 1
Unit Sales:  72892.95826166074
Escenario 2
Unit Sales:  74542.74554463314
```

Figura 23. Predicción de los ingresos en tres escenarios diferentes evaluando el modelo en Python.

Basándonos en las predicciones usando el modelo de regresión lineal estimado, las ventas mensuales de unidades de consolas son más altas en el tercer escenario. Cuando la consola tiene un precio de 8,10 dólares, las ventas mensuales previstas en este escenario son de 74.542 unidades, si redondeamos la respuesta al número entero más bajo.

De esta situación es interesante observar que la opción que da el máximo de ventas de unidades de consolas puede no ser la mejor opción si la compañía quiere maximizar los ingresos, donde los ingresos son las ventas de unidades multiplicadas por el precio de venta. Veamos cual es el escenario que maximiza los ingresos de la compañía de juguetes.

Escenario	Price ($)	Adexp ('000$)	Promexp ('000$)	Unit Sales	Revenue
1	9,10	52,00	61,00	72587,31092	$ 660.544,53
2	7,10	48,00	57,00	72892,95826	$ 517.540,00
3	8,10	50,00	60,00	74542,74554	$ 603.796,24

Figura 24. Predicción de los ingresos en tres escenarios diferentes evaluando el modelo en Excel.

Como se observa en la Figura 24, si el objetivo es maximizar los ingresos de la compañía, el mejor escenario a elegir es el numero uno a pesar de que es el escenario en que se venden menor cantidad de unidades de juguetes. Con esta situación se quiere ilustrar la importancia que tiene definir adecuadamente el objetivo ya que de ello dependerá una adecuada interpretación y análisis de los datos.

El error del modelo.

La regresión es un proceso que tiene errores. ¿Qué significa esta afirmación?, para verlo más claro consideremos de nuevo nuestro modelo de regresión de ventas de consolas. Primero se debe simplificar el modelo de la regresión para entender mejor este concepto de error.

Nuestra variable dependiente de Y sigue siendo las ventas mensuales de unidades de la consola. En esta ocasión la variable independientes o X, será únicamente la

variable - el precio de la consola. Así que nuestro modelo de regresión es como se muestra en las ecuaciones 4 y 5.

$$\boldsymbol{Y} = \boldsymbol{\beta}_0 + \boldsymbol{\beta}_1 * \boldsymbol{X}_1 \qquad (4)$$

$$\boldsymbol{Ventas} = -25096{,}83 - 5055{,}27 * \boldsymbol{Precio} \qquad (5)$$

Este modelo de regresión intenta explicar la variación en las ventas de la consola usando solo el precio como variable explicativa. Rara vez se sabrá cual es la variable descriptiva que realmente influirá en el comportamiento de nuestra variable de interés. Así que para construir un modelo no queda otra opción que asumir una relación funcional, como se ha hecho para el modelo anterior. De esta forma podemos obtener algunos datos de muestra y estimar los parámetros del modelo. Basándonos en el modelo de regresión estimado, podemos calcular las diferencias entre los valores reales de Y observados, y los valores predichos de nuestro modelo de regresión. Estas diferencias se llaman residuales. En nuestro caso, sería la diferencia entre el valor de venta real de nuestros datos y el valor de venta predicho por el modelo de regresión para ese precio en particular. Para ver mejor esto se puede graficar los datos.

$$\boldsymbol{Residuales} = \boldsymbol{Valor}_{actual} - \boldsymbol{Valor}_{\boldsymbol{predicción}} \qquad (6)$$

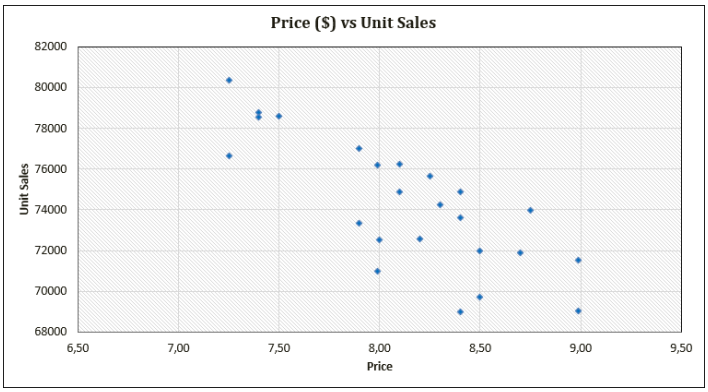


Figura 25. Precio vs Unidades vendidas (Gráfica construida en Excel).

Figura 26. Precio vs Unidades vendidas (Gráfica construida en Python).

Como se observa en las Figura 25 y Figura 26, la relación entre el precio y las unidades vendidas tiene un comportamiento lineal. Además, como se esperaría, se evidencia una tendencia descendente en la que podemos apreciar que a medida que disminuye el precio aumenta el número de unidades que se venden. Esto se debe a que típicamente un aumento en el precio resulta en menores ventas.

Para seguir ilustrando mejor este concepto calcularemos un nuevo modelo usando la metodología anteriormente ilustrada para esta situación particular. En este modelo solo intervendrán las unidades de venta y el precio. Eso implica que calcularemos el coeficiente tomando únicamente las columnas de Unit Sales y Price $.

Excel

Resumen

Estadísticas de la regresión	
Coeficiente de correlación múltiple	0,786759321
Coeficiente de determinación R^2	0,618990229
R^2 ajustado	0,601671603
Error típico	1997,152694
Observaciones	24

ANÁLISIS DE VARIANZA

	Grados de libertad	*Suma de cuadrados*	*Promedio de los cuadrados*	*F*	*Valor crítico de F*
Regresión	1	142558429,5	142558429,5	35,74130137	5,12507E-06
Residuos	22	87749615,41	3988618,882		
Total	23	230308045			

	Coeficientes	*Error típico*	*Estadístico t*	*Probabilidad*	*Inferior 95%*	*Superior 95%*	*Inferior 95,0%*	*Superior 95,0%*
Intercepción	114215,0811	6695,943799	17,05735361	3,60279E-14	100328,5436	128101,6186	100328,5436	128101,6186
Price ($)	-4913,726924	821,9129751	-5,978402912	5,12507E-06	-6618,270107	-3209,18374	-6618,270107	-3209,18374

Figura 27. Resumen de análisis de regresión de Excel.

Python

```
# – – – – – – – – – – – – – – – – – – Evaluación del modelo # 2 – – – – – – – – – – – – – – – – – – –
# Evaluación del modelo Simplificado
unit_v_s1 = []
residual = []
for i in range(0, len(unit_v)):
    unts = Inter_1 + B11 * price[i]
    unit_v_s1. append(unts)
    rsd = unit_v[i] – unit_v_s1[i]
    residual. append(rsd)
# – – – – – – – – – – – – – – – – Imprimir resultado del modelo # 2 – – – – – – – – – – – – – – – –
print (" – – – – Modelo # 2 – – – – – ")
# Parametros
print ('Error : ', error_1 )
print ('Coficiente R^2 : ', R2_1)
print ('Coficiente : ', B11)
```

```
print ('Intercepto : ', Inter_1)
#Ecuación del modelo
print ('Ecuación General de modelo: \n Y = {0} {1} * Price'.format(Inter_1, B11))
print (" - - - - Evaluación / modelo # 2 - - - - - ")
print (" Predicte Sales       Residual")
for i in range(0, len(unit_v)):
   print ("{0} {1} ".format(unit_v_s1[i], residual[i]))
# - - - - - - - - - - - - - - - - - - - - - - - - - - - Gráficas 1 - - - - - - - - - - - - - - - - - - - - - - - - - - -
fig_4 = plt.figure()
plt.plot(price, unit_v, "o", label = "Data")
plt.plot(price, Y_p1, "ro", label = "Prediction")
plt.xlabel("Price")
plt.ylabel("Unit Sales")
plt.title("Unit Sales vs Price")
plt.legend()
fig_4.show()
```

```
 ---- Modelo # 2 -----
Error : 1912.128127397623
Coficiente R^2 : 0.6189902292387627
Coficiente : -4913.7269237283335
Intercepto : 114215.0811014509
Ecuación General de modelo:
 Y = 114215.0811014509 -4913.7269237283335*Price
```

Figura 28. Resumen de análisis de regresión de Python.

De acuerdo con el coeficiente calculado en la Figura 27 y Figura 28, la ecuación del modelo queda de la siguiente forma:

$$\boldsymbol{Ventas} = 114215{,}0811 - 4913{,}7269 * \boldsymbol{Precio} \qquad (7)$$

De la ecuación 7 se puede observar que el coeficiente B1 es negativo y además que es diferente al que se calculó en el modelo anterior. De igual forma ocurre con el B0. Este comportamiento es natural ya que cada vez que se añaden o se quitan variables en el modelo de regresión, las estimaciones de los coeficientes se desplazan y se reajustan. Sin embargo, esto no significa que la relación sea perfecta ya que si calculamos las unidades vendidas usando esta ecuación y algún precio para el que ya conocemos las unidades vendidas de acuerdo con la lista de datos, no obtendremos el mismo valor. Para ver mejor esto se evaluarán los siguientes precios que ya tiene un valor actual (real) dentro de la lista.

Precio (actual)	Unit sales (actual)
$ 8,75	73.959

Tabla 4. Valores reales de la lista de datos de venta de juguetes.

Usamos el valor de precio de la Tabla 4 y evaluamos nuestro modelo.

Estimación	Price ($)	Unit Sales
1	8,75	71.220

Figura 29. Predicción de la venta de unidades de juguetes evaluando el modelo de la ecuación 7 en Excel.

Como se puede apreciar en la Figura 29, que la predicción hecha por el modelo da como resultado 71.220 unidades vendidas. Este valor no corresponde con el número de unidades vendidas que es 73.959, valor que se conocía de los datos originales. Si calculamos la diferencia entre estos dos números habremos calculado el error residual de los datos. Para ello utilizamos la ecuación 8 y la aplicaremos a todos los datos en el archivo Excel y Python, como se observa en la Figura 30 y Figura 31.

$$\boldsymbol{Residuales} = Unit\ Sales - Predicted\ Sales \qquad (8)$$

Month	Unit Sales	Price ($)	Predicted Sales	Residual
1	73959	8,75	71220	2739
2	71544	8,99	70041	1503
3	78587	7,50	77362	1225
4	80364	7,25	78591	1773
5	78771	7,40	77854	917
6	71986	8,50	72448	-462
7	74885	8,40	72940	1945
8	73345	7,90	75397	-2052
9	76659	7,25	78591	-1932
10	71880	8,70	71466	414
11	73598	8,40	72940	658
12	74893	8,10	74414	479
13	69003	8,40	72940	-3937
14	78542	7,40	77854	688
15	72543	8,00	74905	-2362
16	74247	8,30	73431	816
17	76253	8,10	74414	1839
18	72582	8,20	73923	-1341
19	69022	8,99	70041	-1019
20	76200	7,99	74954	1246
21	69701	8,50	72448	-2747
22	77005	7,90	75397	1608
23	70987	7,99	74954	-3967
24	75643	8,25	73677	1966

Figura 30. Cálculo de residuales para el modelo de la ecuación 5 (Excel).

```
---- Evaluación / modelo # 2 -----
 Predicte Sales          Residual
71219.97051882798     2739.029481172023
70040.67605713318     1503.3239428668167
77362.12917348839     1224.8708265116147
78590.56090442048     1773.4390955795243
77853.50186586122     917.4981341387756
72448.40224976005     -462.40224976005265
72939.77494213289     1945.2250578671083
75396.63840399706     -2051.638403997058
78590.56090442048     -1931.5609044204757
71465.65686501439     414.34313498561096
72939.77494213289     658.2250578671083
74413.8930192514     479.10698074860557
72939.77494213289     -3936.7749421328917
77853.50186586122     688.4981341387756
74905.26571162423     -2362.2657116242335
73431.14763450573     815.8523654942692
74413.8930192514     1839.1069807486056
73922.52032687856     -1340.5203268785554
70040.67605713318     -1018.6760571331833
74954.40298086152     1245.597019138484
72448.40224976005     -2747.4022497600527
75396.63840399706     1608.361596002942
74954.40298086152     -3967.402980861516
73676.83398069214     1966.166019307857
```

Figura 31. Cálculo de residuales para el modelo de la ecuación 5 (Python).

Como se puede observar en la Figura 30 y Figura 31, ninguno valor residual es igual a cero, eso implica que ningún valor la predicción coincidió exactamente con el valor real. Podemos ver esta desviación de los valores a través de una gráfica de dispersión (ver Figura 32 y Figura 33).

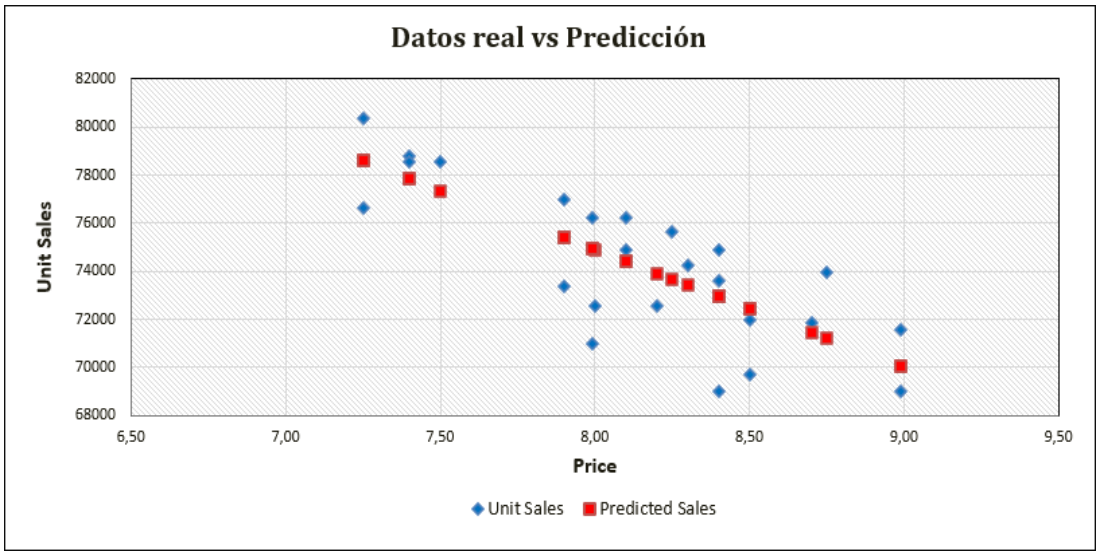


Figura 32. Predicciones vs Datos reales (Excel).

Figura 33. Predicciones vs Datos reales (Python).

Como se observa en la Figura 32 y *Figura 33*, los cuadros en rojo pertenecen a la recta que describe el modelo calculado en la ecuación 5. De esto se puede concluir que, entra más pequeñas sean las desviaciones o los residuos, mejor será el ajuste del modelo. Las desviaciones por encima de la línea de regresión son residuos positivos, mientras que las desviaciones por debajo de la línea de regresión son residuos negativos. La estadística R-cuadrada que se reporta en el informe de la regresión en Excel y Python, es una medida de la bondad del ajuste del modelo de regresión a los datos y se basa en esta noción de errores y su magnitud. El R-cuadrado varía de cero a uno. El valor específico del R-cuadrado se interpreta como la proporción de la variación de la variable Y, que se explica por el modelo de regresión. Y nuestra regresión de las ventas de consolas, R-cuadrado es igual a 0,61899 como se aprecia en la Figura 27 y Figura 28 - esto implica, que este modelo de regresión es capaz de explicar alrededor de 61,9% de variación o cambios en las ventas unitarias del juguete. ¿Qué sucede con el porcentaje restante?, no tiene explicación.

Concepto R^2

El valor de R^2 mide el porcentaje de variación de los valores residuales respecto a la línea que representa el modelo. A través de este valor que varía entre 0 y 1 se mide que tan bien representa el modelo el conjunto de datos.

Se puede notar además respecto al primer modelo que al aumentar el númerode variables X, aumenta R-cuadrado. Es decir, el R-cuadrado era más alto en el modelo cuando también incluimos los gastos de publicidad y promoción. Un mayor valor de R-cuadrado, es decir, más cercano a uno, implica que una mayor proporción de la variación de la variable Y, se explica por el modelo de regresión. Esto implica que el modelo se ajusta bien a los datos. Un menor valor de R-cuadrado, es decir, más cercano a cero, implica que una menor proporciónde

variación en la variable Y, se explica por el modelo de regresión. En otras palabras, el modelo no se ajusta bien a los datos.

No existe un valor de R-cuadrado, por encima del cual se pueda afirmar que se tiene un modelo de buen ajuste, y por debajo del cual se pueda afirmar que se tiene un modelo de mal ajuste.

Se aclara que el mismo concepto se aplica a las situaciones en las que se tiene más de una variable explicativa o X.

Origen del error en la regresión.

- Se omiten variables descriptivas.
- Relación de baja linealidad entre la variable explicativa y la variablede interés.

Supuesto de la regresión lineal.

Los supuestos que permiten aplicar la regresión lineal a situaciones por ejemplo empresariales son los siguientes:

- El error tiene una distribución normal con la media igual a 0.
- La desviación estándar es constante.

Visualmente, lo que estas suposiciones significan es que las barras rojas de error verticales que se muestran en la gráfica de dispersión de la Figura 34 que corresponde al modelo anterior, tienden a estar distribuidas aproximadamente por igual por encima y por debajo de la línea de regresión. De modo que el promedio a través de los errores positivos y negativos tienden a ser aproximadamente 0. Además, la distribución de estas barras verticales de error tiende a ser similares a través de toda la línea de regresión.

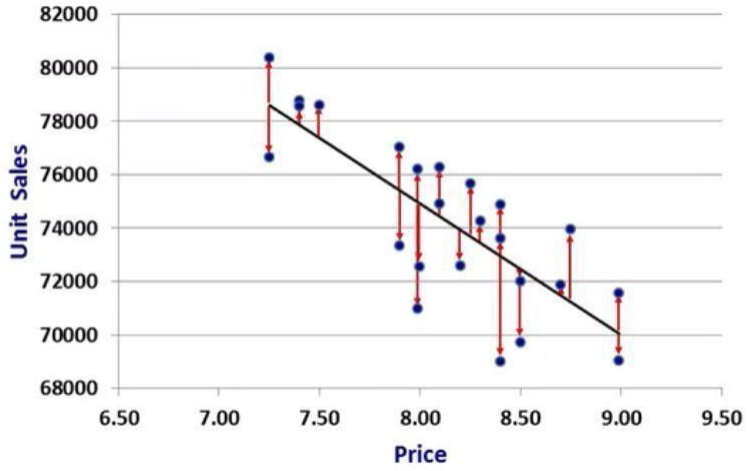


Figura 34. Gráfica de predicción con barrar de error.

Otra forma de pensar sobre esto es que, si se traza un histograma de todos los términos de error utilizando los datos, se tendería a obtener una curva en forma de campana centrada en 0.

La importancia de suposición de normalidad de estos errores se puede ver de la siguiente forma. Se supone un modelo como el mostrado en la ecuación 3.

$$\boldsymbol{Y = \beta_0 + \beta_1 * X_1 + \beta_2 * X_2 + \beta_3 * X_3 + \beta_k * X_K} \qquad (3)$$

Luego obtenemos una muestra de datos y obtenemos estimaciones de los coeficientes beta. Denotemos estas estimaciones que obtenemos para los coeficientes beta como bs. Entonces el modelo se convierte de la siguiente forma.

$$\boldsymbol{Y = b_0 + b_1 * X_1 + b_2 * X_2 + b_3 * X_3 + b_k * X_K} \qquad (9)$$

La relación entre los bs y beta es que los bs son una estimación de la beta. Dependiendo del uso de la muestra para la estimación, el valor de bs puede cambiar. Por ejemplo, en una regresión de ventas de juguetes, si hubiéramos utilizado 36 meses de datos en lugar de 24 meses, podríamos haber reunido estimaciones ligeramente diferentes del impacto del precio y otras variables en las ventas. Esto indica que los propios bs pueden considerarse como variables aleatorias, y a su vez tienen una distribución que es una distribución normal centrada en el denominado valor verdadero de las betas. Por ejemplo, b0 sigue una distribución normal centrada en el valor verdadero de B0. b1 sigue una distribución normal centrada en el valor verdadero de B1, y así sucesivamente.

El comportamiento de las betas y las b es análogo a la relación entre la media de la población y la media de la muestra. La media de la población es fija pero desconocida. Y la media de la muestra puede ser pensada como una variable aleatoria con una distribución normal centrada en la media de la población. De forma similar, en el contexto de la regresión, el valor de las betas es fijo y desconocido. Sin embargo, obtenemos estimaciones de muestra de estas betas fijas desconocidas usando nuestros coeficientes de regresión estimados, que llamamos bs. Estos bs pueden ser pensados como variables aleatorias que tienen una distribución normal centrada en el verdadero valor de las betas.

Para que el trabajo sea significativo y se llegue a resultados importantes se deben llevar a cabo pruebas hipótesis en el contexto de regresión. Para eso nos podemos apoyar en la ecuación 10.

$$\frac{\boldsymbol{b_0 - \beta_0}}{\boldsymbol{s_{b_0}}} \sim \boldsymbol{t}_{n-k-1} \qquad (10)$$

Donde:

$\boldsymbol{b_0}$: Es el valor de B0 que se obtuvo del modelo.
$\boldsymbol{\beta_0}$: Es el verdadero valor de B0 el cual es desconocido.
$\boldsymbol{n}$: Numero de observaciones.
$\boldsymbol{k}$: numero de variables X.

$\boldsymbol{n-k-1}$: Grados de libertad residuales de la ecuación.
$\boldsymbol{S_{b_0}}$: Error estándar de bo que se genera en la salida de la regresión junto a la estimación del coeficiente.

Esta ecuación es la misma para cada una de las betas que conformen el modelo de regresión lineal. Estos resultados nos ayudarán a probar la estabilidad y precisión de nuestros coeficientes estimados. También podremos probar varias afirmaciones sobre la relación de las variables explicativas con nuestra variable de interés, la variable Y.

Como conclusión de lo anterior, se resalta que el verdadero valor de beta nunca se conoce. Todo lo que obtenemos es una estimación de ese coeficiente beta usando una muestra. Sin embargo, podemos probar la precisión de esta usando hipótesis.

5. Proyecto: desarrollo en Python de una herramienta basada en inteligencia artificial que permite la gestión de un repositorio de información.

La dirección de servicios para innovación de CIDET ha recopilado a lo largo de años de operación un volumen de información digital considerable. Dentro de esta información están una gran cantidad de proyectos desarrollados para el sector eléctrico en Colombia. Esta información almacenada es de gran valor para la organización ya que puede emplease para desarrollar proyectos actuales con temáticas similares o puede emplearse como base informativa para ofrecer información de interés para el sector eléctrico en particular y con transformar esa base de datos en un servicio informativo. Sin embargo, este repositorio de información también se puede emplear de forma interna a la organización para evaluar de acuerdo con su historia cuales han sido los tipos de proyectos en los que la organización se ha enfocado en los últimos años, de modo que pueda encontrar allí oportunidades para fortalecer que ofrece servicios y actividades actualmente o que dentro de su plan de expansión se integrarán en un futuro. Una de las dificultades para acceder de forma eficiente y adecuada a esta información se debe a:

- No se tiene conocimiento de todos los temas que hay en la base de datos, ya que los proyectos no fueron ejecutados por el mismo grupo de trabajo. Esto se debe a que es un repositorio de más de 10 años por que las personas que participaron en algunos proyectos ya no se encuentran en la organización y en caso contrario es poco probable que una persona recuerde todo el trabajo de un periodo de tiempo tan largo.

- La información no se guardó respectando algún concepto de clasificación que permitiera comprender de forma global las temáticas o proyectos que se han ejecutado.

Con el fin de brindar una solución al contexto anteriormente mencionado, se desarrollará una herramienta que permita el manejo de grandes volúmenes de información digital (texto digital) de forma que se facilite la obtención, clasificación y empleo de información para todo el personal de servicios para la innovación CIDET. La herramienta propuesta se basa en técnicas de analítica de datos e inteligencia artificial, de forma particular en el procesamiento del lenguaje natural (PLN). La selección de este concepto en particular se debe a la gran aplicabilidad y grandes resultados que ha obtenido actualmente este tipo de técnicas.

El propósito de la herramienta de gestión documental desarrollada se limitará inicialmente a ser capaz tomar un texto con una temática en particular y posteriormente tener la capacidad de determinar el tema central de este, de modo que se pueda clasificar. Este proceso inicial tiene como objetivo ayudar a CIDET a estructurar el repositorio de información y dejarlo preparado para implementación de otras técnicas a través de la cuales se puedan identificar tendencias en los datos. A través de esta cualidad la herramienta facilitará el guardado de información y la disposición posterior de la misma.

A continuación, se mostrará los conceptos de analítica de datos en los que se basa la herramienta de gestión documental, el proceso de desarrollo de la misma y prueba de funcionamiento.

Conceptos

El procesamiento del lenguaje natural (PLN) es un subcampo de inteligencia artificial que se ocupa de la comprensión y el procesamiento del lenguaje humano. A la luz de los nuevos avances en el aprendizaje automático, muchas organizaciones han comenzado a aplicar el procesamiento del lenguaje natural para la traducción, chatbots, filtrado de candidatos, organización documental, clasificadores normativos etc. (Maklin, 2020).

Calculo cuanto se parecen dos documentos:

La tabla muestra un banco de palabras que se seleccionó para comparar dos documentos. Esto se hace mediante (n) que es número de veces que una palabra aparece dentro de cada texto (ver Tabla 5).

Tabla 5. Evaluación de similitud de textos.

Docs.	Palabra 1	Palabra 2	Palabra 3	Palabra 4
Documento 1	n_{11}	n_{12}	n_{13}	n_{14}
Documento 2	n_{21}	n_{22}	n_{23}	n_{24}

La similitud de los textos se calcula como:

$$(n_{11} x\, n_{21}) + (n_{12} x n_{22}) + (n_{13} x n_{23}) + (n_{14} x n_{24}) = \mathbf{S} \qquad (1)$$

Si S es muy grande indica una mayor similitud entre los documentos.

El problema de la longitud de los documentos:

Puede ocurrir que los documentos 1 y 2 sean el doble de largos, esto implicaría que la evaluación de similitud S va a ser mayor, es decir, el tamaño de los textos altera la ponderación. Para resolver este problema, se normalizan los vectores, como ejemplo se normalizará el vector del documento 1 de la siguiente forma:

$$\boldsymbol{V_1} = \sqrt{n_{11}^2 + n_{12}^2 + n_{13}^2 + n_{14}^2} \qquad (2)$$

$$\frac{n_{11}}{\boldsymbol{V_1}}; \frac{n_{12}}{\boldsymbol{V_1}}; \frac{n_{13}}{\boldsymbol{V_1}}; \frac{n_{14}}{\boldsymbol{V_1}} \qquad (3)$$

Este procedimiento se debe hacer para cada uno de los vectores pertenecientes a cada texto. De esta forma se garantiza que S no se verá afectada por el tamaño del texto y que la ponderación será correcta.

Como se prioriza las palabras que realmente nos indican la temática de un texto:

"Decimos que las palabras importantes son aquellas localmente comunes y globalmente raras (Sk, 2020)"

Hay palabras que se repetirán mucho en todos los documentos (verbos y sustantivos muy comunes, preposiciones, etc.) mientras que otras se repetirán mucho únicamente para aquellos documentos relacionados. Las primeras serán palabras comunes en el conjunto de todos los documentos que harán que la medida de similitud entre dos documentos sea más alta aun no estando relacionados. Las segundas son las que realmente nos interesa ponderar, pues son la que realmente diferencian documentos parecidos de documentos sin relación. Es decir, queremos ponderar aquellas palabras que:

- Aparecen mucho en un documento.
- Aparecen poco en el conjunto de todos los documentos.

Para resolver este problema se introduce el siguiente termino vector de frecuencias de términos y frecuencia inversa en documentos o en inglés term frequency - inverse document frequency (tf-idf, por la sigla en inglés).

Para calcular ti-idf debemos partir del vector original donde ya teníamos la frecuencia de cada palabra en un documento. Consideramos también un vector para las mismas palabras donde cada valor se calcula como:

$$\mathrm{IDF\_n_{nn}} = \mathrm{Log}\left(\frac{\text{Número de documentos}}{1+\text{Númerodedocumentosdondeaparecelapalabra}}\right) \qquad (4)$$

Si miramos con detalle esta fórmula para cada palabra, veremos que para aquellas palabras que aparecen en casi todos los documentos el resultado será el logaritmo de un número cercano a 1, el cual se aproxima a cero, mientras que, para palabras que aparecen en pocos documentos, será el logaritmo de un número cada vez mayor conforme la palabra aparece en menos documentos (más rara es) (Sk, 2020).

Luego, para determinar el peso de cada palabra (n) en el texto (p) usamos la siguiente expresión:

$$\text{(IDF D. T., 2020) } \mathbf{TF-IDF} = (\text{TF}) \text{ x } (\text{IDF}_{n_{nn}}) \qquad (5)$$

Donde:

TF-IDF: Peso de una termino (n) en un documento (p).

TF =n_{11}: Frecuencia de un término en documento.

$\text{IDF}_{n_{nn}}$ = Factor de participación de una palabra en el banco de documentos comparados.

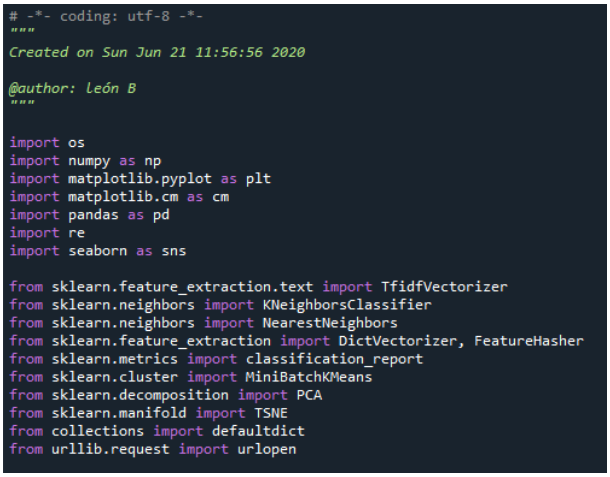
Existen varios métodos para calcular la densidad de palabras presentes en un documento a través del método TF IDF. La ecuación 6, presenta una modificación del IDF, la diferencia con el número apoderado calculado con la ecuación 4 cambia en la quinta cifra decimal lo que para nuestro caso de estudio no será relevante. A continuación, se muestra la ecuación (IDF S. T., 2020).

$$\text{IDF}_{n_{nn}} = \text{Log } \left(\frac{\text{Número de documentos}+1}{1+\text{Número de veces que aparece la palabra}+1}\right) + 1 \qquad (6)$$

Desarrollo

En la Figura 35, se muestran las librerías empleadas para el desarrollo de la herramienta de gestión documental. Las librerías empleadas se distribuyen en los siguientes grupos:

- Visualización de datos y personalización de gráficas: Matplotlib y seaborn.
- Acceso a funciones del ordenador: OS.
- Librarías para la lecturade datos en otros formatos como Excel, csv, txt etc.: Pandas.
- Librarías para la navegación y descargue de información de páginas web: Url y re.
- Librarías para implementación de metodologías de inteligencia artificial y analítica de datos: SkLearn.



```
# -*- coding: utf-8 -*-
"""
Created on Sun Jun 21 11:56:56 2020

@author: León B
"""

import os
import numpy as np
import matplotlib.pyplot as plt
import matplotlib.cm as cm
import pandas as pd
import re
import seaborn as sns

from sklearn.feature_extraction.text import TfidfVectorizer
from sklearn.neighbors import KNeighborsClassifier
from sklearn.neighbors import NearestNeighbors
from sklearn.feature_extraction import DictVectorizer, FeatureHasher
from sklearn.metrics import classification_report
from sklearn.cluster import MiniBatchKMeans
from sklearn.decomposition import PCA
from sklearn.manifold import TSNE
from collections import defaultdict
from urllib.request import urlopen
```

Figura 35. Librerías importadas.

La metodología se basa en la recolección de una base de datos muestral con la que se va a comprar cada uno de los textos que deseemos clasificar. Esta comparación se desarrolla con base a la metodología tf-idf. Para el construir estas bases de datos se eligieron 3 temas aleatorios que se caracterizan por ser muy generales, por lo que existe mucha información relacionada conellos, estos temas son: ciencia, los deportes y economía. Luego se procede a formar una base de datos de textos por temática, como se muestra en la Figura 36. Esta base dedatos contiene archivos txt (0001.data) para cada una los temas, estos archivos contienen pequeños fragmentos de textos tomados de páginas web o blogs de internet y que están relacionados con las tres temáticas seleccionadas. Luego los diferentes archivos son divididos de forma proporcional para crear las carpetas example_set_#, dentro de las cuales se encuentra la misma división por temas del repositorio general. Por último 3 grupos de estos son tomados para realizar el entrenamiento del algoritmo y 1 grupo se deja a parte para realizar una prueba (test) de rendimiento de este. Es de aclarar que la diferencia de número de archivos (000#.data) no genera ningún sesgo dentro del algoritmos ya que los valores de textos son normalizados como se explicó anteriormente en la sección de conceptos.

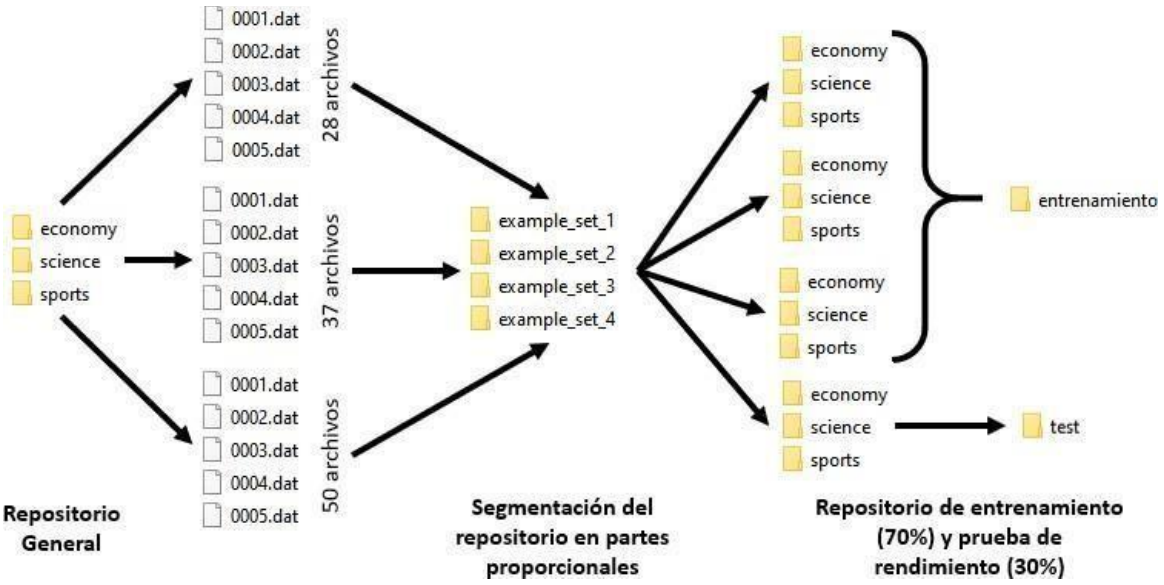


Figura 36. Base de datos.

En la Figura 37, se muestra la lectura de datos, con te proceso se busca almacenar en tres listar todas las palabras de cada uno de los textos guardados en la carpeta "entrenamiento". Estas listas serán la base de comparación para cada uno de los textos que se desee clasificar.

```
#----Leer todos los documentos de un directorio y sus subdirectorios

def read_all_documents(root):
    labels = []
    docs = []
    for r, dirs, files in os.walk(root):
        for file in files:
            with open(os.path.join(r, file), "r", encoding='utf-8') as f:
                docs.append(f.read())
            labels.append(r.replace(root, ''))
    return dict([('docs', docs), ('labels', labels)])


data = read_all_documents('entrenamiento')
documents = data['docs']
labels = data['labels']
```

Figura 37. Lectura de datos.

Luego generar las listas de entrenamiento del algoritmo se procede a calcula la frecuencia de las palabras para una de las listas, esto se logra a través de un sistema de puntaje que cuenta el número de veces que se repite una palabra dentro de cada uno de los textos seleccionados para cada tema, como se muestra en la Figura 38.

```
#----Contar la frecuencia de las palabras en los documentos

def tokens(doc):
    return (tok.lower() for tok in re.findall(r"\w+", doc))

def frequency(tokens):
    f = defaultdict(int)
    for token in tokens:
        f[token] += 1
    return f


def tokens_frequency(doc):
    return frequency(tokens(doc))
```

Figura 38. Frecuencia de las palabras.

A continuación, en el proceso se usan las listas con las frecuencias de las palabras para aplicar el método "vectorizer" de Python con el que se generará una matriz nxn con la calificación tf-idf. Esta matriz es la que contiene todos los parámetros de calificación de las palabras de cada tema y será la plantilla en la que se evaluará cada uno de los textos, como se observa en la Figura 39.

```
#----Extraer las características de los documentos
#-----Nombres de los rasgos simbólicos

vectorizer = DictVectorizer()
vectorizer.fit_transform(tokens_frequency(d) for d in documents)
vectorizer.get_feature_names()
```

Figura 39. Matriz de clasificación tf-idf.

Esta matriz se puede depurar para obtener mejores resultados, para lograr esto lo que se hace es eliminar todas aquellas palabras que distorsionan la calificación de matricial, como lo son: las preposiciones, artículos, adverbios, conectores y verbos auxiliares y preposiciones de modo. Este conjunto de palabras si bien son necesarias para dar sentido a las oraciones no indican el tema de un texto. La eliminación de estas palabras de la matriz de valoración se muestra la Figura40.

```
#-----Entrenar un clasificador de texto usando la agrupación de K-Means
#-----Palabras que se eliminaran de la evaluación

prepositions =['a','ante','bajo','cabe','con','contra','de','desde','en','entre','hacia','hasta','para','por',
               'según','sin','so','sobre','tras']
prep_alike = ['durante','mediante','excepto','salvo','incluso','más','menos']
adverbs = ['no','si','sí']
articles = ['el','la','los','las','un','una','unos','unas','este','esta','estos','estas','aquel','aquella',
            'aquellos','aquellas']
aux_verbs = ['he','has','ha','hemos','habéis','han','había','habías','habíamos','habíais','habían']

tfid = TfidfVectorizer(stop_words=prepositions+prep_alike+adverbs+articles+aux_verbs)

X_train = tfid.fit_transform(documents)
y_train = labels

clf = KNeighborsClassifier(n_neighbors=10)
clf.fit(X_train, y_train)
```

Figura 40. Eliminación de palabras.

Con el finde valorar el desempaño del repositorio de información en conjunto con el algoritmo de predicción, ahora se crea una lista de prueba. Esta lista de prueba

todos los archivos contenidos en la carpeta test, la cual se había mencionado anteriormente. Los artículos contenidos en esta capeta son desconocidos por el algoritmo, es decir, no están dentro de su banco de entrenamiento por lo tanto el algoritmo se verá sesgado en dar un rendimiento de coincidencias del 100%. Para definir el rendimiento de cada banco de temáticas se creó un Score Test independiente para uno, como se observa en la Figura 41.

```
#-----Predecir las categorías de los nuevos artículos

test = read_all_documents('test')
X_test = tfid.transform(test['docs'])
y_test = test['labels']
y_pred = clf.predict(X_test)

print('accuracy score %0.3f' % clf.score(X_test, y_test))

#-----Matriz de métricas - confusión
# confs_matrix = pd.crosstab(y_test , y_pred)
# sns.heatmap(confs_matrix, annot=True, cbar=True)

y_test = list(y_test)
X_test = list(X_test)

science = []
economy = []
sport = []

for i in y_test:
    if i =='\\sports':
        sport.append(i)
    elif i =='\\economy':
        economy.append(i)
    elif i =='\\science':
        science.append(i)

Doc = ['science','economy','sport']
dit = [len(science),len(economy),len(sport)]

print(classification_report(y_test, y_pred))
```

Figura 41. Test de rendimiento del algoritmo.

El resultado del Score Test se puede aprecia en la Figura 42, allí se puede apreciar que la valoración promedio de las comparaciones para cada texto alcanzó puntuación en la exactitud de 94,7%. Lo que es bastante bueno considerando que el repositorio de información más grande solo contiene 50 muestras de texto. De manera particular se puede apreciar que las diferentes temáticas obtuvieron puntajes de exactitud muy similares entre 94% y 97%. Se resalta que la precisión de la temática deportes es inferior debido a que la franja de valores que debe evaluar es un poco mayor debido a que tiene más recursos de comparación. Esto indica que un repositorio de 50 muestras de texto comienza a ser bueno para alcanzar un buen rendimiento sostenido del algoritmo, sin embargo, para alcanzar una robustez considerable es conveniente un base de datos de aprendizaje mucho mayor.

```
accuracy score 0.947
              precision    recall  f1-score   support

    \economy       1.00      0.94      0.97        31
    \science       1.00      0.88      0.94        34
     \sports       0.89      1.00      0.94        48

    accuracy                           0.95       113
   macro avg       0.96      0.94      0.95       113
weighted avg       0.95      0.95      0.95       113
```

Figura 42. Rendimiento del algoritmo.

A continuación, para visualizar mejor el desempeño del algoritmo se crea una base muestral para realizar el cálculo del número óptimo de componentes, es decir,

verificar que la información esté agrupada en 3 clúster, como se muestra en la Figura 43.

```
#-----Cluster con K - means Optimos

def find_optimal_clusters(data, max_k):
    iters = range(2, max_k+1, 2)

    sse = []
    for k in iters:
        sse.append(MiniBatchKMeans(n_clusters=k, init_size=1024, batch_size=2048, random_state=20).fit(data).inertia_)
        print('Fit {} clusters'.format(k))

    f, ax = plt.subplots(1, 1)
    ax.plot(iters, sse, marker='o')
    ax.set_xlabel('Cluster Centers')
    ax.set_xticks(iters)
    ax.set_xticklabels(iters)
    ax.set_ylabel('SSE')
    ax.set_title('SSE by Cluster Center Plot')

find_optimal_clusters(X_train, 20)

clusters = MiniBatchKMeans(n_clusters=4, init_size=1024, batch_size=2048, random_state=20).fit_predict(vectors)
```

Figura 43. Clusterización de la base de datos.

En la Figura 44 , se puede apreciar la clusterización de la base de datos, además parecen las palabras más representativas de cada grupo. Como se puede apreciar las palabras de la base de datos global son bastante similares, esto indica que las listas base de cada tema aún está lo suficientemente caracterizada para reconocer textos en los que no esté bien definida la temática a través de las palabras.

```
Cluster 1
dada,cometido,aisladas,diputados,martti,sonado,equilibrar6n,deflació
n,salvarseaporta,puerto

Cluster 2
gigantesco,perdieron,cometido,llamado,equilibrar6n,aisladas,sonado,s
alvarseaporta,deflación,puerto

Cluster 3
clasificado,luis,lamentando,llamado,equilibrar6n,aisladas,sonado,sal
varseaporta,deflación,puerto
```

Figura 44. Clúster de la base de datos.

Concepto - número Optimo de componente principales: dada una matriz de datos de dimensiones *n* x *p*, el número de componentes principales que se pueden calcular es como máximo de *n*-1 o *p* (el menor de los dos valores es el limitante). Sin embargo, siendo el objetivo del PCA reducir la dimensionalidad, suelen ser de interés utilizar el número mínimo de componentes que resultan suficientes para explicar los datos. No existe una respuesta o método único que permita identificar cual es el número óptimo de componentes principales a utilizar. Una forma de proceder muy extendida consiste en evaluar la proporción de varianza explicada acumulada y seleccionar el número de componentes mínimo a partir del cual el incremento deja de ser sustancial.

Con el fin de identificar gráficamente que tan bien agrupados están los diferentes temas y contrastar lo anteriormente dicho respecto a que los

grupos de temas aún no están bien definidos se usará dos técnicas de clusterización: la técnica t-SNE y la técnica PCA, como se observa en la Figura 45.

```
#-----Plot cluster con K - means

def plot_tsne_pca(data, labels):
    max_label = max(labels)
    max_items = np.random.choice(range(data.shape[0]), size=3000)

    pca = PCA(n_components=2).fit_transform(data[max_items,:].todense())
    tsne = TSNE().fit_transform(PCA(n_components=50).fit_transform(data[max_items,:].todense()))


    idx = np.random.choice(range(pca.shape[0]), size=3000)
    label_subset = labels[max_items]
    label_subset = [cm.hsv(i/max_label) for i in label_subset[idx]]

    f, ax = plt.subplots(1, 2, figsize=(14, 6))

    ax[0].scatter(pca[idx, 0], pca[idx, 1], c=label_subset)
    ax[0].set_title('PCA Cluster Plot')

    ax[1].scatter(tsne[idx, 0], tsne[idx, 1], c=label_subset)
    ax[1].set_title('TSNE Cluster Plot')

plot_tsne_pca(X_train, clusters)
```

Figura 45. Técnica t-SNE y PCA.

Concepto - técnica de clusterización PCA: es una técnica de visualización de datos y reducción de dimensionalidad lineal no supervisada para datos de muy alta dimensión. Dado que tener datos de alta dimensión es muy difícil de obtener información agregando a eso, es muy intensivo en computación. La idea principal detrás de esta técnica es reducir la dimensionalidad de los datos que están altamente correlacionados al transformar el conjunto original de vectores en un nuevo conjunto que se conoce como **componente principal**.
PCA intenta preservar la estructura global de los datos, es decir, cuando convierte datos d-dimensionales en datos d-dimensionales, luego intenta mapear todos los clústeres como un todo debido a que las estructuras locales pueden perderse. La aplicación de esta técnica incluye filtrado de ruido, extracciones de características, predicciones del mercado de valores y análisis de datos genéticos. Este proceso se no hace esto usando conjeturas sino usando matemáticas duras y usa algo conocido como eigenvalores y eigenvectores de la matriz de datos. Estos eigenvectores de la matriz de covarianza tienen la propiedad de que apuntan a lo largo de las principales direcciones de variación de los datos. Estas son las direcciones de máxima variación en un conjunto de datos.

Técnica de clusterización T-SNE: t-SNE también es una técnica de visualización de datos y reducción de dimensionalidad no lineal no supervisada. La matemática detrás de t-SNE es bastante compleja, pero la idea es simple. Incrusta los puntos de una dimensión superior a una dimensión inferior tratando de preservar la vecindad de ese punto.

A diferencia de PCA, intenta preservar la estructura local de datos minimizando la divergencia Kullback-Leibler (divergencia KL) entre las dos distribuciones con respecto a las ubicaciones de los puntos en el mapa. Esta técnica encuentra aplicación en la investigación de seguridad informática, análisis de música, investigación del cáncer, bioinformática y procesamiento de señales biomédicas.

La incrustación de vecinos estocásticos distribuidos en t (t-SNE) es otra técnica para la reducción de la dimensionalidad y es particularmente adecuada para la visualización de conjuntos de datos de alta dimensión. A diferencia de la PCA, no es una técnica matemática sino probabilística. El documento original describe el funcionamiento de t-SNE como:

"La incrustación de vecinos estocásticos distribuidos en t (t-SNE) minimiza la divergencia entre dos distribuciones: una distribución que mide las similitudes por pares de los objetos de entrada y una distribución que mide las similitudes por parejas de los puntos de baja dimensión correspondientes en la incrustación".

Básicamente, lo que esto significa es que analiza los datos originales que se ingresan en el algoritmo y busca la mejor forma de representar estos datos utilizando menos dimensiones al hacer coincidir ambas distribuciones. La forma en que lo hace es computacionalmente bastante pesada y, por lo tanto, existen algunas limitaciones (serias) para el uso de esta técnica. Por ejemplo, una de las recomendaciones es que, en caso de datos dimensionales muy altos, es posible que deba aplicar otra técnica de reducción de dimensionalidad antes de usar t-SNE (Toward data science, 2016).

En la Figura 46, se puede apreciar el resultado de la aplicación de la técnica de clusterización t-SNE y PCA. Podemos notar que ambas técnicas pudieron identificar los tres clústeres (temáticas), sin embargo, como se había mencionado anteriormente las temáticas no están bien distinguidas. Esto indica que la clasificación correcta de un texto depende fuertemente de que el texto a evaluar esté muy bien definido en término de palabras, de lo contrario la clasificación será errónea. Este inconveniente se soluciona incorporando una base de datos de entrenamiento más robusta.

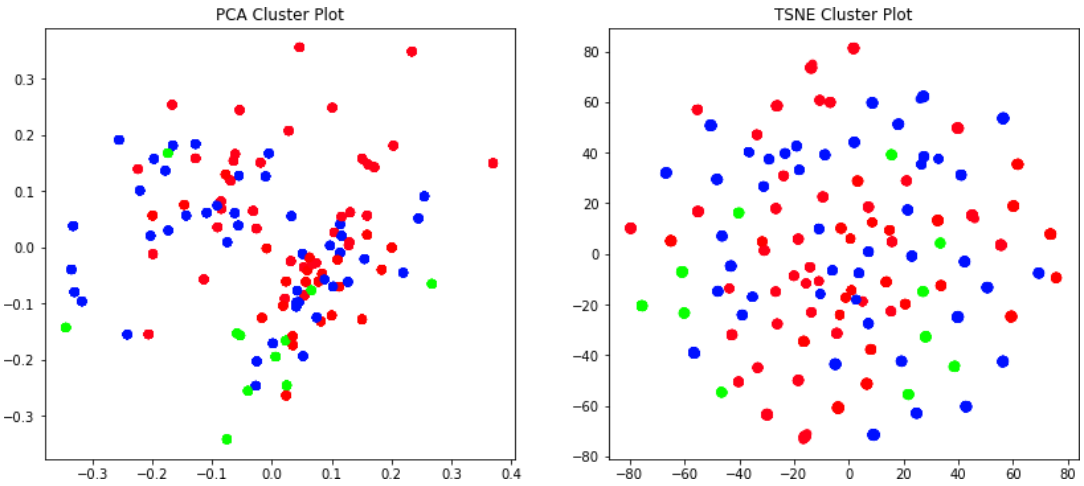


Figura 46. Comparación de las técnicas de Clúster visual t-SNE y PCA.

Cuando la matriz tf-idf está totalmente preparada se procede crear la matriz donde se almacenarán los datos de entrada, es decir, la matriz del texto que se quiere evaluar, como se muestra en la Figura 47. Para generar una dinámica rápida se creó un sistema de captura de url, a través de este proceso se toma la url de una página web (la que se desee) y se ingresa como input al algoritmo.

```
#-----Ingreso de nuevos artículos

def predict_category(url, classifier):
    article = urlopen(url).read()
    X_test = tfid.transform([article])
    return clf.predict(X_test)[0]

def show_predicted_categories(urls, classifier):
    for url in urls:
        print('predicted category: ' + predict_category(url, clf))

doc_clas = []
Ingres = str(input('Desea Clasificar un documento (si/no):'))


while Ingres == 'si':
    doc_clas_in = doc_clas.append(str(input('Ingrese la URL del artículo:')))
    show_predicted_categories(doc_clas,clf)
    Ingres = str(input('Desea Clasificar un documento (si/no):'))
```

Figura 47. Esquema de datos de entrada.

De esta forma se descarga toda la información de la página web a una lista de palabras la cual se vectoriza y se compara con la matriz tf-idf de cada uno de los temas, determinando así el puntaje para cada una y definiendo la finalmente la temática del texto que se evaluó.

A continuación, se muestra un ejemplo a través del cual se pone a prueba el algoritmo en la clasificación de un texto. Para la prueba se selecciona la siguiente url: "https://www.futbolred.com/colombianos-en-el-exterior/james-rodriguez-hoy-tiempo-sin-ser-titular-en-real-madrid-jugara-este-domingo-liga-de-espana-2020-119194". Esta página web reporta noticias relacionadas con el mundo del futbol

por lo que en todos sus árboles de código existen etiquetas y texto estrechamente relacionado con la temática.

Figura 48. Ejemplo prueba para la herramienta de clasificación.

Para evaluar el texto, el algoritmo inicialmente pregunta si desea clasificar un documento, posteriormente pide al usuario ingresar la url de la página web que desea evaluar, como se muestra en la Figura 49.

```
Desea Clasificar un documento (si/no):si

Ingrese la URL del artículo:https://www.futbolred.com/colombianos-en-el-exterior/james-rodriguez-hoy-
tiempo-sin-ser-titular-en-real-madrid-jugara-este-domingo-liga-de-espana-2020-119194
```

Figura 49. Ingreso de datos de prueba.

Luego de correr el algoritmo, como salida este indica la temática a la que pertenece el texto de la url indicada, como se muestra en la Figura 50.

```
Ingrese la URL del artículo:https://www.futbolred.com/colombianos-en-el-exterior/james-rodriguez-hoy-
tiempo-sin-ser-titular-en-real-madrid-jugara-este-domingo-liga-de-espana-2020-119194
predicted category: \sports
```

Figura 50. Salida: herramienta de clasificación.

Se a validado la correcta operación del algoritmo de acuerdo con los conceptos de clasificación tf-idf. Se resalta que el algoritmo se puede someter a muchas mejoras en términos de visualización, rendimiento y adaptación de más características como la valoración de la información histórica almacenada. Para su implementación en la gestión del repositorio de información se debe realizar una recopilación extensiva de documentos que puedan servir como base de datos de aprendizaje para el algoritmo. A través de esta herramienta se puede lograr una clasificación rápida del extenso repositorio del área de servicios para innovación CIDET.

Conclusiones

- La apropiación de las técnicas de vigilancia tecnológica ofrece las herramientas a través de las cuales se puede crear un proceso estandarizadoy altamente selectivo a través de la cual se captura información de fuentes estructuradas y no estructuradas y posteriormente analizarla, depurarla y seleccionarla con base en las necesidades y requerimientos del sector, empresa o personas. Luego de esta etapa de consolidación de la información se pasa a la etapa de comunicación, esta etapa es crucial, ya que es donde se pone en evidencia para el cliente los hitos más importantes de la vigilancia y el verdadero valor en la información. Esta última etapa es crítica ya que de su planeación depende en gran medida el éxito en la transferencia del valor en la información, por lo que es importante la consolidación de habilidades de comunicación efectiva.

- La vigilancia tecnológica se puede enfocar de diferentes formas de acuerdo con las necesidades de una organización en particular, esto se refiere a que la vigilancia puede enfocarse en aspectos completamente diferentes tales como: 1) detección y reducción de riesgos, con lo que se busca identificar factores que puedan influir de forma negativa a las organizaciones. En este apartado se puede incluir, por ejemplo, los estudios de mercado y de apropiación tecnológica. 2) Prospectiva o anticipación, esto permitirá a las empresas puedan anticiparen en temas relacionados con la tecnología, competidores, productos, servicios, identificando así tendencias o información de valor estratégico para las organizaciones. 3) comparación, con este aspecto se busca conocer las debilidades y fortalezas de los competidores y de esta forma poder comprar el negocio. 4) cooperación o colaboración, permite identificar las oportunidades de cooperación con socios, comunidades, movimientos sociales etc. Para detectar oportunidades de colaboración. Con este proceso se puede abrir nuevas oportunidades de mercado, y a su vez, facilita la integración de nuevas tecnologías, productos, procesos o la expansión sectorial de la organización. 5) Innovación, posibilita la identificación de oportunidades de mejora y desarrollo de ideas innovadoras en un entorno de mercado, investigativo o social, por lo que ayuda a definir la estrategia del+D+i.

- La vigilancia tecnológica se basa en la recopilación de datos de diferente tipo y de diferentes fuentes de información, en su mayoría estos datos son digitales por lo que pueden ser sometidos a múltiples transformaciones, como por ejemplo codificaciones, este proceso permite que la información se pueda medir, agrupar, clasificar etc. Por esta razón, la vigilancia tecnológica en la

actualidad está adoptando todas las técnicas de analítica de datos, Big Data y la inteligencia artificial para desarrollar sus metodologías propiamente estandarizadas. De esta forma podrá recopilar el gran flujo de información digital e incorporar valor en la misma. Por otra parte, esto representa un gran desafío para las empresas en términos de perfil profesional, ya que el perfil de los vigías exige: un conocimiento en algún área en particular (por ejemplo, una carrera profesional), habilidades en búsqueda y además se tenga habilidades en técnicas de analítica de datos.

- La incorporación de tecnología, plataformas, técnicas, métodos de inteligencia artificial, ciencia de datos y Big Data en los procesos empresariales, les brinda a las empresas la oportunidad de optimizar sus actividades, identificar oportunidades, disminuye la incertidumbre en determinadas situaciones lo que facilita la toma de decisiones y es crucial para fortalecer a las organizaciones haciéndolas altamente competitivas en esta época de transformación digital.

Referencias Bibliográficas

ACQUIA. (15 de 4 de 2020). *CQUIA*. Obtenido de CQUIA: https://www.acquia.com/resources/whitepaper/acquia-earns-top-score-wcm-strategy-forrester-wave?cid=7013a000002KXMdAAO&ct=search&ls=google&lls=pro_latam_transformation_ovrdrv_se&gclid=EAIaIQobChMIt6Km4brH6AIVU9yGCh1kXwJcEAAYBCAAEgKDnPD_BwE&gclsrc=aw.ds

Aguirre J.J., A. A. (2012). Unidad de Inteligencia Estratégica Tecnológica: Vigilancia tecnológica e inteligencia competitiva para el sector eléctrico colombiano. Primera edición. . En A. A. Aguirre J.J., *Unidad de Inteligencia Estratégica Tecnológica: Vigilancia tecnológica e inteligencia competitiva para el sector eléctrico colombiano. Primera edición.* (págs. ISBN: 978-958-57193-2-3.). Medellín: L. Vieco e Hijas Ltda.

Aldasoro Alustiza, J. C. (July 18-20, 2012.). La vigilancia tecnológica y la inteligencia competitiva en los estándares de la calidad en I+D+i. *6th International Conference on Industria Engineering and Industrial Management. XVI Congreso de ingería de Organización.* Vigo.

ASOCODIS. (01 de 07 de 2020). *ASOCODIS.* Obtenido de ASOCODIS: http://www.asocodis.org.co/index.php?option=com_archivos&view=archivos&id=64&Itemid=11

ATRI. (06 de 10 de 2020). *ATRI.* Obtenido de ATRI: https://truckingresearch.org/wp-content/uploads/2017/10/ATRI-Operational-Costs-of-Trucking-2017-10-2017.pdf

BBVA. (15 de 4 de 2020). *www.bbva.com*. Obtenido de www.bbva.com: https://www.bbva.com/es/que-es-el-regtech/

El Tiempo. (02 de 07 de 2020). *El Tiempo.* Obtenido de El Tiempo: https://www.eltiempo.com/economia/asi-crecera-la-economia-de-colombia-en-el-2020-segun-banco-mundial-449866#:~:text=Econom%C3%ADa-,Crecimiento%20de%20Colombia%20seguir%C3%A1%20acelerando%20este%20a%C3%B1o%3A%20Banco%20Mundial,impulso%20de%20la%20actividad

ESFI. (04 de 2020). *ESFI.* Obtenido de ESFI: https://www.esfi.org/workplace-injury-and-fatality-statistics

ESFI-3D. (04 de 2020). *ESFI-3D.* Obtenido de ESFI-3D: https://www.esfi.org/home-disaster-safety

Forrester. (15 de 4 de 2020). *Forrester*. Obtenido de Forrester: https://www.forrester.com/search?N=0+21089&sort=3&everything=true&dateRange=1&searchOption=0

G.D., V. (febrero de 2013 p.). Investigación, Desarrollo e innovación Empresarial. . *ResearchGate: Technnical Report – Universidad Federal de Integrañao Latino*.

Galileo Technologies. (23 de 06 de 2020). *Galileo Technologies.* Obtenido de Galileo Technologies: https://www.galileoar.com/gas-versus-diesel-transporte-mas-sustentable/

Gartner. (15 de 4 de 2020). *Gartner*. Obtenido de Gartner: https://www.gartner.com/doc/reprints?id=1-1YDUKTC6&ct=200217&st=sb

Gartner. (2020). *Gartner.* Obtenido de Gartner: www.gartner.com/doc

GOV.CO. (2020). *GOV.CO.* Obtenido de GOV.CO: https://www.datos.gov.co/Minas-y-Energ-a/Superservicios-Informaci-n-de-Accidentes-de-Origen/es62-3x6p

IDF, D. T. (4 de 04 de 2020). *USEO.* Obtenido de USEO: https://useo.es/tf-idf-relevancia/

IDF, S. T. (20 de 04 de 2020). *Sklearn TF IDF.* Obtenido de Sklearn TF IDF: https://www.youtube.com/watch?v=iioBLaRgMMQ

IEA. (11 de 06 de 2020). *IEA.* Obtenido de IEA: https://webstore.iea.org/download/direct/288

Innovation Origins. (2020). *Innovation Origins.* Obtenido de Innovation Origins: https://innovationorigins.com/electric-trucks-economically-and-environmentally-desirable-but-misunderstood/

IVECO - Natural Power. (11 de 06 de 2020). *onthemosway.* Obtenido de onthemosway: https://www.onthemosway.eu/wp-content/uploads/2015/06/TrainMoss-II-Madrid-26-11-2015.pdf

Keraunos. (23 de 06 de 2020). *http://apolo.creg.gov.co/.* Obtenido de http://apolo.creg.gov.co/: http://apolo.creg.gov.co/Publicac.nsf/52188526a7290f8505256eee0072eba7/c55a6288b2a5fd1a05257cfb0054ae53/$FILE/Circular036-2014%20Anexo.pdf

LNG UE. (16 de 06 de 2020). *SEVENTH FRAMEWORK PROGRAMME.* Obtenido de SEVENTH FRAMEWORK PROGRAMME:

http://lngbc.eu/system/files/deliverable_attachments/LNG_BC_D%203%208%20Cost%20analysis%20of%20LNG%20refuelling%20stations.pdf

Maklin, C. (24 de 04 de 2020). *Towars Data Science.* Obtenido de Towars Data Science: https://towardsdatascience.com/natural-language-processing-feature-engineering-using-tf-idf-e8b9d00e7e76

Ordoñez, C. P. (30 de 06 de 2020). *http://bdigital.unal.edu.co/.* Obtenido de http://bdigital.unal.edu.co/: http://bdigital.unal.edu.co/54166/7/1126420820.2016.pdf

Procolombia. (1 de 07 de 2020). *Procolombia.* Obtenido de Prooclombia: https://procolombia.co/noticias/covid-19/coronavirus-y-su-impacto-en-la-economia-colombiana

ResearchGate. (11 de 06 de 2020). *Energías .* Obtenido de Energías : https://www.researchgate.net/publication/330921964_Fuel_Switch_to_LNG_in_Heavy_Truck_Traffic

Rev Dinero. (1 de 07 de 2020). *Rev Dinero.* Obtenido de Rev Dinero: https://www.dinero.com/economia/articulo/como-va-la-demanda-de-energia-en-colombia-durante-2020/287112

Sanchez, J. L. (– ALTEC 2009). Redes de unidades de vigilancia tecnológica e inteligencia competitiva (vtic). *Artículo presentado en el XIII Seminario Latino-Iberoamericano de Gestión Tecnológica*, (pág. 10). Cartagena.

Sk, S. (20 de 05 de 2020). *Software Sk.* Obtenido de Software Sk: https://www.joragupra.com/2016/03/clasificacion-automatica-de-textos.html

SUI. (01 de 07 de 2020). *SUI.* Obtenido de SUI: http://www.sui.gov.co/web/energia/reportes/tecnico-operativo/informacion-del-formato-8

Toward data science. (2016). *Toward data science.* Obtenido de Toward data science: https://towardsdatascience.com/visualising-high-dimensional-datasets-using-pca-and-t-sne-in-python-8ef87e7915b

UPME. (1 de 07 de 2020). *UPME.* Obtenido de UPME: http://www.siel.gov.co/siel/documentos/documentacion/Demanda/Proyeccion_Demanda_Energia_Jul_2019.pdf

XM - noticias. (15 de 4 de 2020). *XM noticias*. Obtenido de XM noticias: https://www.xm.com.co/Lists/noticias/DispForm.aspx?ID=1712&ContentTypeId=0x010060BD47A5D614E84E9E3FFDBED2A72A9C00AFB418AFE810B14F8614540433AA0FA9

XM. (15 de 4 de 2020). *XM SA ESP*. Obtenido de XM SA ESP: https://www.youtube.com/watch?v=GFXRz3nwFDk&t=762s

.....