

Aplicación de técnicas Big Data a la predictibilidad del gasto energético.

Proyecto ConSumar

Fundación COMPUTAEX
 info@{computaex.es, cenits.es}
 CénitS – Centro Extremeño de iNvestigación, Innovación Tecnológica y Supercomputación
 Cáceres, Extremadura, España

Resumen—El consumo energético es uno de los principales gastos que deben afrontar, tanto los ciudadanos como las empresas. Las nuevas tecnologías juegan un papel principal a la hora de establecer sistemas de control y seguimiento con vistas a gestionar el consumo y la sostenibilidad energética. La adecuada planificación de los equipos con gran consumo puede llegar a resultar un ahorro importante. El proyecto ConSumar busca el desarrollo de herramientas que, basadas en técnicas de Big Data, permita a los consumidores una mejor planificación del gasto energético a través de la predicción de los costes, relacionando el precio de la energía con los datos climáticos.

Índice de Términos— *gestión de la energía, smart cities, big data, open data, supercomputación, cloud computing*

I. INTRODUCCIÓN

En los últimos años, una de las principales preocupaciones, tanto de los consumidores particulares, como de las empresas son los cambios en el precio de la energía que conllevan una gestión ineficiente económicamente así como medioambientalmente [1].

Es por esta problemática que el principal objetivo de este proyecto es poner al servicio de los consumidores y de las empresas extremeñas un conjunto de herramientas que permitan la gestión del consumo energético, así como una planificación adecuada de los periodos en los que el ahorro por un menor coste energético pueda resultar interesante.

Esto es posible ya que el sistema tarifario eléctrico, que entró en funcionamiento en España en abril de 2014, permite a los consumidores que dispongan en sus domicilios de contadores inteligentes, dotados de discriminación horaria, consumir con mayor precisión, eligiendo tarifas y proveedores con libertad diaria.

Se pretende, por tanto, analizar el nuevo sistema tarifario y todas sus facetas de manera que permita a los consumidores elegir las tarifas más ventajosas en cada momento y también una plataforma que permita conocer las experiencias de otros

consumidores.

En este proyecto se han alcanzado los siguientes objetivos:

- Lograr eficiencia energética, tanto en el ámbito doméstico como empresarial.
- Aplicación de técnicas de Big Data al análisis de los datos de consumo energético.
- Búsqueda de la sostenibilidad económica y ecológica a través de las nuevas tecnologías.
- Un producto software que actúe como prospector de las diversas tarifas eléctricas en cada momento.

Una solución obvia a las problemáticas objeto de estudio es intentar hacer un mejor uso de los recursos existentes, es decir, una utilización inteligente de los recursos energéticos. En este sentido, uno de los principales gastos que tienen afrontar en su día a día, tanto los consumidores, como las empresas son los costes derivados del consumo energético. Si se lograra optimizar la gestión de los recursos energéticos, se mejorarían considerablemente, tanto los costes, como la huella de carbono generada por ciertas formas de producción de la energía.

El propósito de una gestión energética eficiente es la de informar o dirigir a los consumidores a modificar los patrones de conducta de manera que planifiquen mejor las horas en las que pueden conectar los equipos con un gran consumo de energía. Ajustarse y adaptarse a las condiciones energéticas del momento es una tarea muy compleja e inviable sin la gestión de la información con las TIC.

En esta línea de trabajo, toma especial relevancia la utilización de un adecuado sistema de información adaptado a las características del problema a tratar, en donde la gran cantidad de datos que pueden ser usados para la toma de decisiones requieran de un análisis exhaustivo de los mismos, para inferir posibles comportamientos en la red eléctrica y, por tanto, los costes asociados en cada momento del día.

De entre los diferentes factores, son de gran importancia y se han de tener en cuenta los parámetros ambientales, climáticos, horarios, eventos especiales y cualquier otro que pudiera afectar a los patrones de generación y consumo de energía.

Asimismo, es importante contar con las herramientas de cómputo y comunicaciones adecuadas a la tarea, que requerirán una capacidad de procesamiento de información elevada y la utilización de la infraestructura de comunicaciones para que el usuario pueda acceder a la información de manera sencilla e, incluso, integrarla con diferentes equipamientos.

Debido principalmente a las restricciones de tiempo y recursos, se plantea la realización de un piloto del proyecto, en el que se desarrollan algunas de las tecnologías que habrían de conformar el sistema de ayuda en su totalidad, de forma que permitan vislumbrar las capacidades de un sistema de información completo como el que se pretende desarrollar como trabajo futuro del proyecto.

El punto de partida de este proyecto, se centra en la toma de datos en los costes energéticos diarios desde marzo de 2014, momento en el que se implanta la nueva normativa energética. Los datos climáticos también se han tomado desde ese mes para incorporarlos al sistema de Big Data.

En el proyecto ConSumar se ha estudiado la legislación vigente del sector eléctrico, una herramienta de gestión de costes, así como datos relativos a la inferencia entre los datos climáticos y el posible coste energético asociado a los mismos. Finalizado este desarrollo, el objetivo es ampliar el ámbito de estudio a las diferentes localizaciones clave para la generación de energía en España.

La siguiente sección de este artículo presenta el estudio del sector eléctrico español y del sistema tarifario implantado. La sección III describe las diferentes posibilidades para desarrollar una arquitectura Big Data que pueda ser aplicada para la obtención y el análisis de los datos del sector eléctrico. La sección IV presenta la arquitectura desarrollada en el proyecto así como los avances obtenidos para la obtención, procesado y visualización de los datos. Por último, la sección V muestra las conclusiones del trabajo realizado así como las posibles contribuciones futuras que se podrían obtener a través de su continuación.

II. EL SECTOR ELÉCTRICO

Se define el suministro de energía eléctrica como la entrega de energía a través de las redes de transporte y distribución mediante contraprestación económica en las condiciones de regularidad y calidad que resulten exigibles. Las actividades destinadas al suministro de energía eléctrica son: generación, transporte, distribución, servicios de recarga energética, comercialización e intercambios intracomunitarios e internacionales, así como la gestión económica y técnica del sistema eléctrico.

Desde el año 1997, el sector eléctrico en España ha experimentado una profunda transformación. Hasta ese

momento, la actividad del sector eléctrico era llevada a cabo por empresas caracterizadas por una importante estructura vertical, y que ejercían monopolio en las distintas regiones españolas. La liberalización del mercado de energía eléctrica comenzó en el año 1997 teniendo como objetivo adaptarse a las normas europeas sobre suministro eléctrico.

Las características del mercado eléctrico han hecho que éste sea un proceso largo y complejo con un acceso difícil para el consumidor para la contratación de aquellas tarifas más acordes a sus necesidades.

La Ley 54/1997 [2], supuso el inicio del proceso de liberalización progresiva del sector mediante la apertura de las redes a terceros, el establecimiento de un mercado organizado de negociación de la energía y la reducción de la intervención pública en la gestión del sistema.

La ley desarrolla la regulación del sector eléctrico, con el triple objetivo de garantizar el suministro eléctrico, la calidad de dicho suministro y que se realice al menor coste posible. Para ello, no se considera necesario que el Estado se reserve para sí el ejercicio de ninguna de las actividades que integran el suministro eléctrico. La explotación unificada del sistema eléctrico nacional deja de ser un servicio público y sus funciones son asumidas por dos sociedades mercantiles y privadas, que serán responsables de la gestión económica y técnica del sistema. La planificación estatal queda restringida a las instalaciones de transporte.

En cuanto a las tarifas regulatorias de venta del suministro eléctrico, se han producido dos cambios regulatorios muy importantes pero que no varían las opciones de contratación, sino sólo la tarificación. El Real Decreto 1718/2012 [3], y el Real Decreto 216/2014 [4], establecen estos cambios.

En el Real Decreto 1718/2012 se determina el procedimiento para realizar la lectura y facturación de los suministros de energía en baja tensión con potencia contratada no superior a 15 kW. La elevación del umbral de 10 a 15 kW permitirá acogerse a la facturación bimestral de consumo real no sólo a los consumidores domésticos, sino también a las PYMES.

La empresa comercializadora de último recurso efectuará la facturación de los consumidores acogidos a la tarifa de último recurso basándose en lecturas reales. La lectura de la energía será realizada por el encargado de lectura (empresa distribuidora de cada zona) y se pondrá a disposición de la empresa comercializadora de último recurso.

La lectura será mensual o bimestral en función del equipo de medida:

- Equipos de medida analógicos o convencionales: la lectura de la energía se realizará con periodicidad bimestral.

- Equipos de medida con capacidad para teledadida y telegestión y efectivamente integrados en los correspondientes sistemas (equipos digitales o de última generación): la lectura se hará con una periodicidad mensual. Antes del 31 de diciembre del 2018, los contadores de los suministros de potencia contratada de hasta 15 kW deberán ser sustituidos por equipos que permitan la discriminación horaria y la telegestión (DA 1ª Orden ITC/3860/2007).

El Real Decreto 216/2014, establece la metodología de cálculo de los precios voluntarios para el pequeño consumidor de energía eléctrica y su régimen jurídico. En él, se determinará la estructura de los precios voluntarios para el pequeño consumidor (antes tarifas de último recurso), que serán de aplicación a los consumidores de baja tensión con potencia contratada hasta 10 kW. Se fija el procedimiento de cálculo del coste de producción de energía eléctrica que incluirá el precio voluntario para el pequeño consumidor, de tal forma que se respete el principio de suficiencia de ingresos, aditividad y que no ocasionen distorsiones de la competencia en el mercado, posibilitando su revisión.

El cálculo de los precios voluntarios para el pequeño consumidor (PVPC) se determina por una serie de componentes, desde impuestos o alquiler de equipos a precios fijos y variables que dependen de nuestro consumo y características de la instalación. El precio de la electricidad vendrá dado por cuatro componentes distintos, un término de potencia, un término de energía del peaje de acceso, un término correspondiente al coste horario de la energía y, en su caso, un término de la energía reactiva.

El término de energía correspondiente al periodo de facturación correspondiente, expresado en euros, será el sumatorio resultante de multiplicar la energía consumida durante el periodo de facturación en cada periodo tarifario, por el precio del término de energía correspondiente. El periodo de facturación, generalmente bimensual, será aquel en que se incluirán una serie de días y que se pagarán en una única factura.

Debido a la variabilidad de este valor, será el dato en el que se centrará este estudio. Aunque existen dos formas de cálculo, de acuerdo con el uso de los contadores inteligentes, en este trabajo se ha focalizado el cálculo y análisis de las instalaciones que tengan un contador inteligente. Por lo tanto, para suministros que cuenten con equipos de medida con capacidad para teledadida y telegestión, y efectivamente integrados en los correspondientes sistemas, el valor del coste de facturación de energía activa (FEU) se obtendrá:

$$FEU = \sum [(E_p * TEU) + \sum (E_{ph} * TCUh)]$$

Donde:

- E_p = energía consumida en el periodo tarifario p expresada en kWh.
- E_{ph} = Energía consumida en la hora h del periodo tarifario p , expresada en kWh.
- $TEUp$ = precio del término de energía del precio voluntario para el pequeño consumidor del periodo tarifario p , expresado en euros/kWh.
- $TCUh$: precio del término de coste horario de energía del precio voluntario para el pequeño consumidor, en cada hora h , expresado en euros/kWh.

Se tienen en cuenta, por tanto, el coste del peaje para cada periodo tarifario o día (resultado de multiplicar la energía consumida en el periodo tarifario y el precio del término de energía del precio voluntario para el pequeño consumidor del periodo tarifario p) y el consumo realizado en cada hora que se calcula consultando la energía consumida en la hora h del periodo tarifario p y multiplicándolo por el precio del término de coste horario de energía, en cada hora h .

De entre todos los valores anteriores, el principal valor que variará en la fórmula será el término $TCUh$ cuyo valor evoluciona en función de la hora de consumo. La siguiente fórmula muestra el cálculo de este parámetro:

$$TCUh = (1 + PERDh) * CPh$$

Donde:

- $TCUh$: Término de coste horario de energía del PVPC en cada hora, expresado en euros/kWh.
- CPh : Coste de producción de la energía suministrada en cada hora expresado en euros/kWh.
- $PERDh$: Coeficiente de pérdidas del peaje de acceso de aplicación al suministro en la hora h . Se determina en función del nivel de tensión, del peaje de acceso y, en su caso, del perfil de consumo.

De manera que se puede extraer de la fórmula que el valor que depende de factores externos es la variable CPh . Este valor es calculado mediante la suma de los siguientes conceptos: el precio medio horario obtenido (Pmh); el valor correspondiente a los servicios de ajuste del sistema asociados al suministro (SAh) y otros costes asociados al suministro. Todos estos parámetros son informados por Red Eléctrica Española (REE) para cada hora.

La REE calcula los términos de los que se compone el coste de producción de la energía y los publica en su página web antes de las 20:15 del día anterior al del suministro para cada una de las 24 horas del día siguiente. A estos efectos, el operador del mercado pondrá a disposición del operador del sistema antes de las 20:00 del día anterior los datos necesarios de precios y cantidades resultantes del mercado.

Con todos estos datos se debe construir una plataforma que permita almacenar de manera histórica los valores, así como

predecir el comportamiento del mercado energético futuro.

III. EL PARADIGMA BIG DATA APLICADO AL CONSUMO ENERGÉTICO

En los últimos años, el despliegue de las nuevas tecnologías está llegando a sectores tradicionales como el sector energético definiendo nuevos paradigmas y modelos productivos a los ya existentes desde la revolución industrial. El Internet de la Energía (Internet of Energy, IoE) se define como una infraestructura de red que integra múltiples actores a través de diversos estándares y protocolos de comunicación abiertos. De esta manera, se puede interconectar la red eléctrica con Internet permitiendo que la energía y los servicios asociados a la misma, como la generación, el almacenamiento y la distribución puedan ser gestionadas de manera eficiente, almacenando e intercambiando información relativa a cada servicio [5].

Todos los datos que comienzan a recabar las operadoras de electricidad por medio de los nuevos dispositivos instalados en la red, se ha incrementado la cantidad de información que se conoce de los usos y las tendencias de los consumidores. Por tanto, es necesario introducir tecnologías que permitan gestionar ese volumen de información y darle coherencia entre los distintos tipos de datos que se obtienen.

Uno de los paradigmas que permite la gestión y el análisis de grandes volúmenes de datos es Big Data. Este paradigma puede ser considerado como un ecosistema de aplicaciones y tecnologías que ha ido incrementando su complejidad debido, principalmente, a la proliferación de herramientas relacionadas con Big Data. De manera que no hay una única solución para la construcción de una arquitectura Big Data. Esta arquitectura debe permitir almacenar, procesar, analizar y visualizar datos.

Uno de los elementos clave del paradigma Big Data es el almacenamiento de una gran cantidad de datos. Las ingentes cantidades de datos que cualquier usuario o servicio generan, están creando un conocimiento muy importante, tanto para las empresas, como para los investigadores y los propios consumidores. Es por esto importante analizar las diferentes alternativas a la hora de almacenar los datos en la solución basada en Big Data.

Big Data es además un paradigma que necesita de una gran cantidad de cómputo para el procesamiento y almacenamiento de los datos, de manera que logre transformar un gran volumen de datos de distinto formato y origen en información útil.

Es necesario disponer de una gran cantidad de espacio para el almacenamiento, así como mecanismos que permitan el análisis casi en tiempo real. Además, mucha de la información que se utilizará en los análisis de Big Data vendrá de la propia web, un modelo claramente basado en la computación en la nube.

Aunque las tecnologías para infraestructura incorporan el

análisis (o la analítica) de los datos, hay tecnologías específicas que están diseñadas con capacidades analíticas. Algunas de las categorías de analítica en las que se puede dividir Big Data incluyen: Plataformas analíticas, de visualización, de Business Intelligence y herramientas de Machine Learning.

Para el desarrollo del proyecto ConSumar se han analizado diferentes soluciones buscando la arquitectura que mejor se adapte a las necesidades concretas del mismo.

IV. DESARROLLO DE CONSUMAR

Tras el estudio detallado de la normativa que rige el sector eléctrico y de las necesidades y del ecosistema disponible para aplicar el paradigma de Big Data al sector energético, se ha desarrollado una plataforma abierta para la obtención de los datos y su análisis centrado, principalmente, en el coste de la energía.

Estos datos son analizados por las herramientas de análisis para encontrar patrones de comportamiento entre el coste eléctrico y las diferentes variables independientes a medir, como la información climática, la capacidad para generación de energías renovables, etc.

Para ello, se ha diseñado un conjunto de herramientas que realice las siguientes labores:

- Recogida de datos fijos. Para el análisis se debe disponer tanto de los costes energéticos, como de los datos climáticos.
- Herramienta de visualización de las diferentes tarifas eléctricas.
- Plataforma de satisfacción del servicio eléctrico.
- Herramientas de análisis de datos.

La arquitectura del proyecto es descrita de manera visual en la Figura 1. En las siguientes secciones se realizará un análisis detallado de cada una de las herramientas desarrolladas desplegar la plataforma de Big Data del proyecto ConSumar sobre el supercomputador LUSITANIA [6].

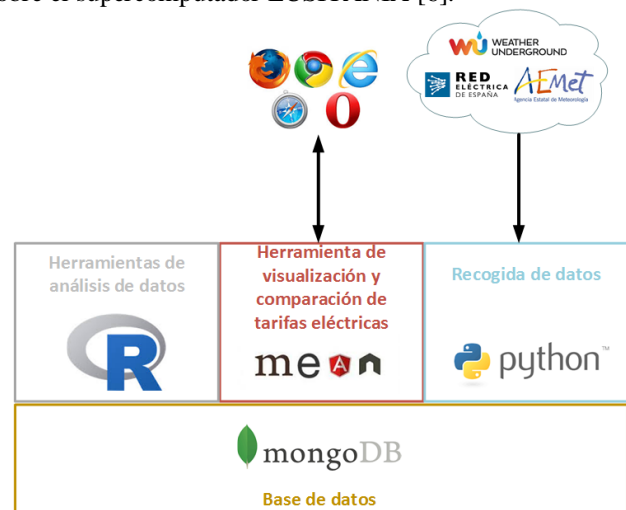


Figura 1: Arquitectura del proyecto ConSumar**A. Herramientas para el almacenamiento de la información no estructurada**

En los últimos tiempos son muchas las aplicaciones y servicios que utilizan el almacenamiento de la información en bases de datos para su correcto funcionamiento. Esas bases de datos eran normalmente SQL como son MySQL, Oracle o MS SQL, pero, desde hace ya algún tiempo, ha aparecido otro paradigma en el mundo de la bases de datos que reciben el nombre de NoSQL que cambia el planteamiento inicial sobre el diseño de las bases de datos relacionales.

Actualmente, la cantidad de información a almacenar, así como la diversidad de los datos, han hecho que los paradigmas tradicionales de las bases de datos relacionales tengan ciertos inconvenientes debidos, principalmente, a problemas de escalabilidad y rendimiento de las bases de datos donde acceden miles de usuarios concurrentes y con millones de consultas diarias.

Además, las bases de datos NoSQL son sistemas de almacenamiento de información que no cumplen con el esquema entidad-relación. Tampoco utilizan una estructura de datos en forma de tabla donde se van almacenando los datos sino que para el almacenamiento hacen uso de otros formatos como clave-valor, mapeo de columnas o grafos.

Estas aproximaciones ofrecen ciertas ventajas sobre las bases de datos relacionales. Entre las más significativas se pueden destacar:

- Se ejecutan en máquinas con pocos recursos: Estos sistemas, a diferencia de los sistemas basados en SQL, no requieren de apenas computación, por lo que se pueden desplegar en equipos de un coste más reducido.
- Escalabilidad horizontal: Para mejorar el rendimiento de estos sistemas simplemente se consigue añadiendo más nodos, con la única operación de indicar al sistema cuáles son los nodos que están disponibles.
- Pueden manejar gran cantidad de datos: Esto es debido a que utiliza una estructura distribuida, en muchos casos mediante tablas Hash.
- No generan cuellos de botella: El principal problema de los sistemas SQL es que necesitan transcribir cada sentencia para poder ser ejecutada, y cada sentencia compleja requiere además de un nivel de ejecución aún más complejo, lo que constituye un punto de entrada en común, que ante muchas peticiones puede ralentizar el sistema.

Para la realización del proyecto se han estudiado las siguientes plataformas: Cassandra, Redis, mongoDB y CouchDB.

Analizadas las diferentes plataformas que se han mencionado anteriormente, se han estudiado las utilidades que proporcionan con respecto al uso y el tipo de datos que se van a almacenar. En el sistema desarrollado, los datos serán un conjunto de documentos de diferente índole para gestionar el conocimiento del sistema de la información. Esta característica conlleva elegir la mejor plataforma que aúne, tanto rápidas consultas de información, así como el almacenamiento de diferentes datos heterogéneos en la plataforma.

Dado que el planteamiento inicial en el proyecto es almacenar datos de los costes energéticos, climáticos y además desarrollar una plataforma que permita la comunicación de la opinión de los consumidores sobre las comercializadoras, se ha propuesto el uso de una base de datos común para todos estos tipos de datos. La aproximación que más se ajusta es la basada en mongoDB por los beneficios de ser una base de datos documentales [7].

B. Herramientas de recogida de datos diarios

En la primera fase del desarrollo se ha tomado la información, tanto de los costes energéticos diarios como de la climatología. Para ello se han desarrollado varias aplicaciones en Python que permiten la recogida de los datos. Las fuentes de los datos de donde se alimenta a la base de datos diseñada en mongoDB son las siguientes:

- Datos relacionados con el precio diario, web de Red Eléctrica de España.
- Datos climáticos históricos, Weather Underground.
- Predicción climática a tres días vista, Agencia estatal de meteorología.

Se han desarrollado tres "parsers" diferentes que gestionan la transformación de los diversos formatos que se extraen de las fuentes de datos originarias a JSON (JavaScript Object Notation) para su posterior almacenamiento en la base de datos:

- Transformación de XML a JSON a través del método `xml2json` usando las librerías `ElementTree` de Python.
- Transformación de CVS a JSON usando el método `csv2json` que se basa en las librerías CVS.
- Por último, la gestión de los datos en hojas de cálculo y su transformación en JSON. Para ello, se ha realizado un script que utiliza los módulos implementados en las librerías `xldr` de Python.

Estos scripts se deberán ejecutar una vez al día para obtener la información. Para ello, se han configurado en el planificador de tareas del servidor. Los horarios de obtención de datos se han programado para obtener la información lo más actualizada posible. Por lo tanto, la obtención de datos se realiza:

- Datos diarios de coste energético a las 23 horas.
- Predicción climática para los próximos 3 días a las 5:50.

- Datos climáticos del día a las 23:55.

C. Herramientas de visualización de tarifas y precios diarios

Para el desarrollo de la herramienta de visualización de tarifas, y dado que la base de datos elegida es mongoDB, se ha optado por un desarrollo basado en la pila de aplicaciones MEAN [8].

Desde los años 90, el lenguaje JavaScript ha incrementado su importancia en el desarrollo de aplicaciones web. Las pilas de desarrollo basado en JavaScript se han ido adaptando de manera que en su comienzo se ejecutaban en el lado cliente (el navegador), pero, actualmente, también está implantándose en el servidor y en la capa de almacenamiento. En cualquiera de esos puntos se puede encontrar módulos de JavaScript para realizar el desarrollo de cualquier aplicación.

Gracias a eso, actualmente es posible crear aplicaciones distribuidas utilizando el mismo lenguaje JavaScript en todas sus fases y capas. A este concepto, a través de las tecnologías que lo posibilitan, se ha denominado al framework como MEAN, acrónimo formado por las iniciales de las cuatro tecnologías principales que entran en juego: MongoDB, Express, AngularJS y Node.js. La Figura 2 muestra la estructura del framework MEAN.

La aplicación está basada en una interfaz en AngularJS que muestra la información en tiempo real de los precios de la energía. Para acceder a esos datos debe hacer una petición al motor API REST de nodeJS y express que harán de controlador de la aplicación (Paso 1 de la figura), gestionando las peticiones y realizando las consultas sobre la base de datos implementada en mongoDB (Paso 2).

Una vez la base de datos devuelve el objeto con la información que ha sido requerida (Paso 3). El objeto devuelto es procesado de nuevo por el API REST y se transforma en la información que deberá mostrar AngularJS a través de HTML (Paso 4). Para ello se construye un objeto JSON que será la información que procesará AngularJS. La Figura 3 muestra la interfaz web de la aplicación desarrollada para el proyecto ConSumar.



Figura 2: Estructura del framework MEAN



Figura 3: Interfaz web de la herramienta de visualización

D. Plataforma de satisfacción del servicio eléctrico

Al igual que la aplicación descrita en la sección anterior, la plataforma de satisfacción del servicio eléctrico suministrará los datos a la base de datos descrita anteriormente, basada en mongoDB.

Para mantener la homogeneidad entre las diferentes herramientas desarrolladas y dado que la base de datos será la misma que se ha expuesto anteriormente, la aplicación estará también basada en el paradigma MEAN.

Esta aplicación tiene dos interfaces diferenciadas. En una primera interfaz se realiza el alta de todas las distribuidoras que se quieran incluir en la plataforma de satisfacción y por las cuales el usuario podrá exponer su grado de satisfacción.

Se debe dar de alta la información de la comercializadora, así como el número de preguntas que se quieran realizar sobre la misma. Una vez se terminan de introducir todas las preguntas, se crea la encuesta y queda almacenada en la base de datos para que los consumidores puedan rellenarla.

La Figura 4 muestra el aspecto de la página para la creación de una nueva comercializadora.

Figura 4: Alta de comercializadora

Cada vez que un usuario accede a la plataforma, la página que se le presenta será genérica mostrando las diferentes encuestas de las operadoras que se encuentren activas en ese momento. La Figura 5 muestra la interfaz inicial de la aplicación de satisfacción.

Toda esta información se incorpora al sistema de Big Data como parte importante de la analítica de datos. La Figura 6 muestra el ejemplo de una posible consulta sobre una distribuidora.

El sistema almacena tres tipos de esquemas diferentes para los datos. El primer esquema de la encuesta almacena una pregunta y un conjunto de opciones. Esas opciones serán el segundo tipo de esquema, en el que se guarda, tanto el texto de la opción, como el número de votos en tiempo real de la opción. El sistema también almacena quién ha expresado su opinión, guardando la dirección IP del votante y el comentario que ha realizado, ya que la aplicación permite recoger, tanto la opinión del usuario en forma de encuesta, como por escrito en un campo de sugerencias. Además, al almacenar la dirección IP del votante, se controla quién ha dado su opinión eliminando votos duplicados de los usuarios.

E. Herramientas de análisis de datos

Una vez expuestas las herramientas para la recolección y visualización de las diferentes fuentes de datos del proyecto, se va a describir la herramienta de análisis de datos para el estudio de los costes de la energía a través de técnicas de Big Data.

Para el análisis de los datos almacenados en mongoDB se han desarrollado diversas aplicaciones usando el lenguaje de programación R bajo el paradigma de Machine Learning [9].

Para cada uno de los conjuntos de datos de las diversas fuentes de datos, se ha creado un Dataframe que incluye los datos almacenados de manera que se acelere el tratamiento de los datos al estar ya en formato R.

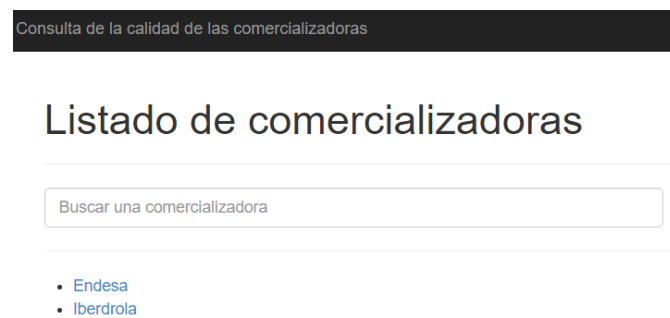


Figura 5: Página principal de aplicación de satisfacción

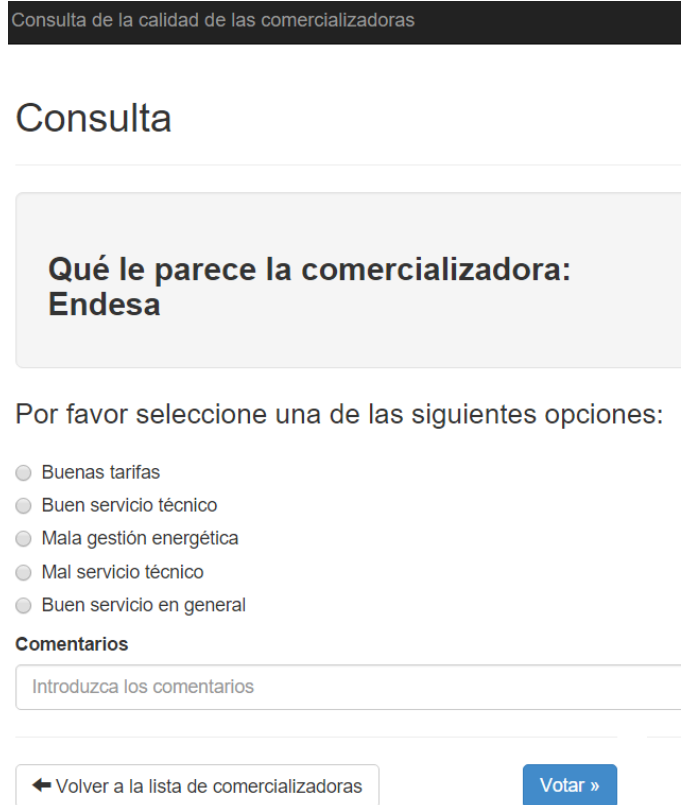


Figura 6: Consulta sobre una de las distribuidoras

Una vez que todos los datos son dispuestos en Dataframes, se analizan relacionándolos entre sí. En este proyecto se han utilizado tres fuentes diferentes de datos abiertos. Por un lado, los costes energéticos desde el año 2014; la información climática desde el mismo año y, por último, las predicciones climatológicas a dos días vista.

Gracias a las fuentes de datos abiertas, los datos de las diferentes arquitecturas están disponibles de forma libre para que cualquier aplicación, usuario o servicio pueda usarlo, sin restricciones de derechos de autor, patentes u otros mecanismos de control. De hecho, este tipo de datos son comparados con otros movimientos y comunidades abiertas como la del Software Libre. Además las fuentes de datos abiertas permiten una fácil transformación entre formatos para que puedan ser reutilizados con mayor facilidad [10].

Con la aplicación de los datos abiertos al Big Data se consigue un conjunto de valores que puedan ser tratados para obtener patrones, realizar predicciones y tomar decisiones sobre la mejor forma de ahorrar en el consumo energético. La Figura 7 muestra una representación de los valores que se han utilizado en este trabajo para el análisis y la toma de decisiones.

Como se puede observar de la Figura 7, la dispersión de los datos es muy grande ya que la variación del coste de producción de la energía denotado por *CPh*, descrito en la Sección 2,

depende de muchos factores, como la producción de energías renovables o más indirectamente el mes o la hora que se tome de referencia para el precio.

Para un análisis más exhaustivo se han realizado diferentes particiones a los datos de manera que se analicen sus interacciones. Se han creado tres grandes particiones dependiendo del tipo de tarifa a la que se le aplique: 2.0A, 2.0DHA o 2.0DHS. También se han realizado particiones por mes y por hora. Con la partición de los datos resultante se realiza una predicción del coste de la energía teniendo en cuenta los parámetros climáticos, la hora y el mes. Para ello se utiliza un modelo de regresión lineal múltiple.

Estos datos son, finalmente, almacenados de nuevo en la base de datos de mongoDB para ser también incorporados, tanto a la aplicación que muestre los resultados, como al propio sistema de Big Data.

Esta aplicación desarrollada para tal efecto bajo el paradigma MEAN dispone los valores horarios de la predicción eléctrica como se muestra en la Figura 8.

V. CONCLUSIONES Y TRABAJO FUTURO

La gestión inteligente de la energía no sólo ayuda a tomar decisiones a los consumidores del momento en el que deben programar el funcionamiento de los equipos que mayor cantidad de energía consumen. Este hecho permite una mayor planificación dependiendo de diferentes factores como los económicos, o, incluso, la implicación con el medio ambiente.

Como se ha visto, las líneas de trabajo en este ámbito son muy amplias y permiten abarcar muchos campos de acción. En este proyecto se han desarrollado un buen número de herramientas que permiten, no sólo la recolección de los datos de fuentes heterogéneas, sino que se clasifican y se disponen para la consulta de manera sencilla a los consumidores.

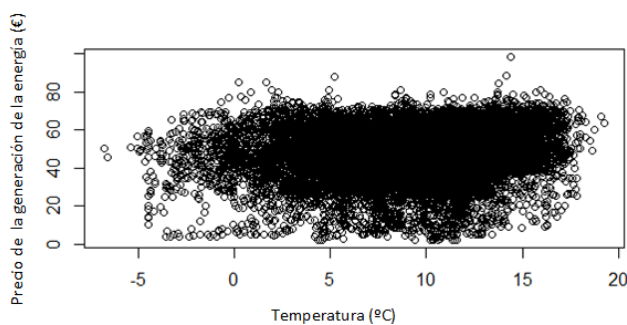


Figura 7: Representación de los puntos de los datos disponibles



Hora	Precio kWh
0 - 1	0.089 €
1 - 2	0.085 €
2 - 3	0.084 €
3 - 4	0.083 €

Figura 8: Datos mostrados de la predicción del sistema de Big Data

Estos datos, además, son tratados para obtener nuevo conocimiento a raíz de la relación de diferentes fuentes de datos como son los datos climáticos. A través de esta información se puede predecir el comportamiento de los mercados de manera que se pueda calcular los posibles costes energéticos y por lo tanto, la planificación de las tareas que mayor consumo energético conlleven.

Este trabajo, por tanto, es el inicio de un sistema inteligente que toma datos de fuentes abiertas, procesa la información, aprendiendo de ella, y, además, puede tomar decisiones teniendo en cuenta predicciones futuras.

Además como trabajos futuros se ha de tener en cuenta que el proyecto se ha tomado los datos abiertos de tres fuentes diferentes, como son el precio de la energía, datos climáticos y predicciones de esos datos para días venideros.

Este conjunto de datos genera una gran cantidad de información para la predicción a través de mecanismos de Machine Learning, pero el número de variables independientes a tener en cuenta es pequeño.

Por tanto, uno de los trabajos principales a desarrollar en futuras etapas del proyecto será la toma y la relación de nuevas fuentes de datos para relacionarlas con las ya existentes, de manera que se incrementen esas variables independientes y el modelo de predicción sea mucho más rico.

Así mismo, es importante integrar el sistema en un entorno de producción donde se puedan tomar decisiones en base a la información que calcula el sistema inteligente. En este caso, las siguientes fases de este proyecto llevarán a la interconexión de la información obtenida por las herramientas desarrolladas en el proyecto con otros sistemas expertos en los que la toma de decisiones basadas en el consumo eléctrico sea un requerimiento.

Este trabajo es también muy importante para centros como CénitS, ya que el consumo eléctrico es uno de los principales gastos para el mantenimiento de los equipos.

El desarrollo de sistemas inteligentes que planifiquen las tareas que mayor consumo energético así como la introducción de la toma de decisión a la hora de introducir fuentes de energía alternativas como la geotermia o la energía fotovoltaica serán, en un futuro no muy lejano, mecanismos para el ahorro de los costes de mantenimiento en el CPD.

Como aplicación directa del proyecto se propone el desarrollo de un sistema de planificación de procesos en un CPD inteligente que base su toma de decisiones en diferentes parámetros, donde la predicción de los costes energéticos sea uno de los principales pilares.

Agradecimientos

Este trabajo de investigación ha sido financiado por el Proyecto CENITAL-2015 incluido en el Programa Operativo FEDER de Extremadura 2007-2013. Eje 1 “Desarrollo de la Economía del Conocimiento (I+D+i, Educación, Sociedad de la Información y TIC). Tema Prioritario 1, Actividades de I+DT en Centros de Investigación Código de Subproyecto: 2014.14.02.007 , Código de Proyecto: 2014.14.02.0010.

REFERENCIAS

[1] Preocupación efectos cambio climático y energía. <http://www.expansion.com/economia/2016/01/18/569d3f0546163f3b098b461d.html>

- [2] BOE. Real Decreto 1718/2012. Boletín Oficial del Estado, 2013, vol. 12, pag. 1704 - 1709. https://www.boe.es/diario_boe/txt.php?id=BOE-A-2013-385
- [3] BOE. Ley 54/1997, del Sector Energético. Boletín Oficial del Estado, 1997, vol. 285, pag. 35097 - 35126. https://www.boe.es/diario_boe/txt.php?id=BOE-A-1997-25340
- [4] BOE, Real Decreto 216/2014. Boletín Oficial del Estado, 2014, vol. 77, pag. 27397 - 27428. <http://www.boe.es/buscar/doc.php?id=BOE-A-2014-3376>
- [5] Huang, Alex Q., et al. The future renewable electric energy delivery and management (FREEDM) system: the energy internet. Proceedings of the IEEE, 2011, vol. 99, no 1, p. 133-148.
- [6] Supercomputador LUSITANIA <http://www.cenits.es/cenits/lusitania/caracteristicas-lusitania>
- [7] MongoDB. <https://www.mongodb.com/es>
- [8] Dickey, Jeff. Write Modern Web Apps with the MEAN Stack: Mongo, Express, AngularJS, and Node.js. Pearson Education, 2014.
- [9] Zhao, Yanchang. R and data mining: Examples and case studies. Academic Press, 2012.
- [10] Open Data. <http://opendata.cloudbcn.cat/MULTI/es/what-is-open-data>