

## Detección de anomalías en el consumo de electricidad para clientes no regulados

### Data Collection

---

#### What data will you collect or create?

Los datos para el proyecto han sido recopilados por el cliente.

Los datos consisten en las lecturas de los instrumentos de medición del consumo de electricidad de los consumidores industriales del mercado no regulado peruano abonados a la compañía ElectroDunas. Contiene datos de identificación de cada cliente: el número de identificación del cliente, el Departamento, la Provincia, el Distrito, la Subestación, las Coordenadas de localización. Además, la información técnica corresponde a los valores históricos de la energía activa entregada (kWh), la energía reactiva entregada (kVarh) y el voltaje reportado cada 15 minutos por el medidor de cada cliente.

El volumen de los datos para el prototipo consiste de decenas de miles de datos, en productivo se espera que la solución cuente con cientos de miles de datos. El formato de los datos son archivos planos de texto con datos separados por algún carácter especial.

#### How will the data be collected or created?

Los datos son recolectados por el cliente y cumplen con los estándares del regulador de servicios públicos del Perú (OSINERGMIN), que le permite a la Compañía soportar la factura al usuario o consumir del servicio.

El proyecto contara con una estructura de carpetas, la cual permitirá organizar los datos de manera jerárquica por vigencia, Departamento y Provincia.

Los archivos de datos será etiquetados de acuerdo al formato Departamento\_provincia\_AAAAMMDD, donde el texto departamento y provincia corresponderán a la ubicación geopolítica del cliente y fecha en formato AAAAMMDD corresponde a la fecha en que fueron capturados los datos en el medidor.

### Documentation and Metadata

---

#### What documentation and metadata will accompany the data?

Los datos se describen claramente a través del nombre. Para esto se crea un diccionario de datos, en cual se incluirá aparte del nombre, una descripción del dato con su naturaleza, rango y un ejemplo del formato del dato.

Los metadatos a considerar y a acompañar la información consisten en anexos de informes de avance donde se documentarán análisis exploratorios de los datos incluidos, así como lo relativo a las metodologías empleadas y resultados de los análisis en cada una de las fases o iteraciones del proyecto.

El título del archivo contendrá las fechas de corte y generación de los datos, así como el nombre del contribuidor

que añade los datos y de quien de los Stakeholders lo obtuvo.

## **Ethics and Legal Compliance**

---

### **How will you manage any ethical issues?**

La organización otorgó el consentimiento de uso de los datos, por el momento su uso está restringido al personal involucrado en el proyecto, para publicar o compartir los datos y o cualquier resultado derivado se tiene que contactar y tener permiso explícito y escrito de la compañía.

Parte de la información que se maneja corresponde a información sensible de cada uno de los clientes de la compañía, por lo cual con objeto de no revelar esta, se reemplazan los números de identificación por una secuencia y se eliminara de los dataset la información sensible como nombres, sexo, teléfono de cada cliente.

### **How will you manage copyright and Intellectual Property Rights (IP/IPR) issues?**

Los datos y el resultado del proyecto son propiedad del cliente. El alcance actual no incluye licencia de reuso de los datos por lo que no se pueden compartir con ningún tercero salvo que se cuente con autorización explícita y escrita por parte del cliente.

## **Storage and Backup**

---

### **How will the data be stored and backed up during the research?**

La etapa de exploración y desarrollo de la solución se adelantará con una muestra de los datos, buscando reducir el impacto de los costos de almacenamiento.

La administración de los archivos del proyecto se realizara en un servidor en la nube, el cual tendrá como herramienta de administración y control DVC(Data Versión Control), sistema que permite el manejo de versiones de los datos y del pre procesamiento realizado sobre estos.

El proveedor del servidor en la nube será Amazon.

El equipo del proyecto, como el cliente, serán los responsables de los respaldos, para lo cual como equipo tendremos backups incrementales del repositorio DVC y de manera semanal se le entregara una copia completa de los datos en su estado actual al cliente para que estos sean a su vez respaldados en su infraestructura.

### **How will you manage access and security?**

Los datos serán almacenados en un servidor en la nube, el cual solo permite el acceso a usuarios que se encuentren registrados y tengan los privilegios suficientes de consulta y escritura sobre el repositorio. La administración del

repositorio se encuentra a cargo de uno de los miembros del equipo, el cual es el responsable de todos los temas de infraestructura y servidores.

El acceso al repositorio estará limitado al equipo del proyecto, debido a que la información es confidencial y sensible.

## **Selection and Preservation**

---

### **Which data are of long-term value and should be retained, shared, and/or preserved?**

Los datos a conservar a largo plazo corresponden a los casos que se identificaron como potenciales pérdidas no técnicas, de modo que se pueda mantener un histórico con la precisión de la herramienta en la detección de anomalías. Esto se realiza con el objeto de reducir el volumen de información a almacenar.

### **What is the long-term preservation plan for the dataset?**

El servicio utilizado para el almacenamiento y control de los datos tiene un costo alto, al hacer parte de los servicios de AWS, razón por la cual una vez finalizado el proceso de investigación, los datos y los resultados de la investigación serán albergados en un proyecto privado en GitHub, herramienta también orientada al manejo y control de versiones, pero la cual es libre y no requiere ningún tipo de inversión o costo para su mantenimiento.

## **Data Sharing**

---

### **How will you share the data?**

El alcance inicial del proyecto es privado por lo que no se contempla compartir el proyecto más allá de los propósitos contractuales en los que se enmarca el inicio del proyecto. En este caso se compartirá el acceso al repositorio correspondiente. Por su bajo volumen, las solicitudes de acceso se manejarán directamente y de forma manual.

Los datos del cliente solo estarán a disposición del equipo del proyecto durante el tiempo que dure este. Posteriormente, toda la información quedará bajo el dominio y custodia del cliente.

### **Are any restrictions on data sharing required?**

Los datos y los resultados de la investigación podrán ser accedidos por el equipo del proyecto y los miembros de la organización que determine el cliente.

## **Responsibilities and Resources**

---

### **Who will be responsible for data management?**

Todo el equipo será responsable por el manejo de datos. La implementación y revisión periódica del DMP será colaborativa.

**What resources will you require to deliver your plan?**

Dependiendo del tipo de implementación que se determine al final, algunos cargos por uso de maquinas virtuales y almacenamiento podrían surgir.

---

## Planned Research Outputs

### Interactive resource - "Visualización de consumos históricos por cliente"

Recurso interactivo de visualización de consumos históricos por cliente. Donde se puede seleccionar el id de cliente, las fechas de interés y como retorno tener un objeto interactivo donde se pueden visualizar los consumos históricos.

### Audiovisual - "Informe periodico de anomalías encontradas"

Reporte de clientes y periodos correspondientes a anomalías encontradas en los consumos de energía.

### Data paper - "Informe de metodologías y resultados."

Informe de metodologías usadas y resultados principales en detección de anomalías.

---

## Planned research output details

Title	Type	Anticipated release date	Initial access level	Intended repository(ies)	Anticipated file size	License	Metadata standard(s)	May contain sensitive data?	May contain PII?
Visualización de consumos históricos por cliente	Interactive resource	Unspecified	Restricted	None specified		None specified	None specified	No	No
Informe periodico de anomalías encontradas	Audiovisual	Unspecified	Open	None specified		None specified	None specified	No	No
Informe de metodologías y resultados.	Data paper	Unspecified	Open	None specified		None specified	None specified	No	No