**ESCUELA POLITÉCNICA NACIONAL**

**FACULTAD DE INGENIERÍA EN SISTEMAS**

**Sistema de minería de datos del portal web SETEC para análisis de oferta y demanda al definir indicadores estáticos y dinámicos**

**NOMBRE DEL COMPONENTE**

**INGENIERO CIENCIAS DE LA COMPUTACIÓN**

**SERGIO ANDRÉS JIMÉNEZ REINO**

**DIRECTOR: JULIÁN GALINDO**

**DMQ, marzo 2023**

CERTIFICACIONES

Yo, Sergio Andrés Jiménez Reino declaro que el trabajo de integración curricular aquí descrito es de mi autoría; que no ha sido previamente presentado para ningún grado o calificación profesional; y, que he consultado las referencias bibliográficas que se incluyen en este documento.

**  
Sergio Jiménez**

Certifico que el presente trabajo de integración curricular fue desarrollado por Sergio Jiménez, bajo mi supervisión.

**  
Julián Galindo  
DIRECTOR**

Certificamos que revisamos el presente trabajo de integración curricular.

**  
NOMBRE\_REVISOR1  
REVISOR1 DEL TRABAJO DE INTEGRACIÓN CURRICULAR**

**  
NOMBRE\_REVISOR2  
REVISOR2 DEL TRABAJO DE INTEGRACIÓN CURRICULAR**

DECLARACIÓN DE AUTORÍA

A través de la presente declaración, afirmamos que el trabajo de integración curricular aquí descrito, así como el (los) producto(s) resultante(s) del mismo, son públicos y estarán a disposición de la comunidad a través del repositorio institucional de la Escuela Politécnica Nacional; sin embargo, la titularidad de los derechos patrimoniales nos corresponde a los autores que hemos contribuido en el desarrollo del presente trabajo; observando para el efecto las disposiciones establecidas por el órgano competente en propiedad intelectual, la normativa interna y demás normas.

Sergio Jiménez

Julián Galindo

NOMBRE\_COLABORADOR(ES)

DEDICATORIA

El presente documento dirigida a mi familia quien me ha apoyado incondicionalmente en los aspectos de mi vida académica.

AGRADECIMIENTO

A mis padres Sergio Jiménez y Bibiana Reino por todo lo que han hecho por mí, me han enseñado y me han hecho la persona que soy hoy en día.

A mis hermanos Juan José y María de Lourdes de quienes he aprendido e inspirado en ser mejor cada día.

A profesores y compañeros quienes me enseñaron, inspiraron y apoyaron en distintos momentos de la carrera universitaria.

Al profesor Roberto Andrade, Denys Flores, Andrés Merino, Sang Wo, Henry Paz y Julián Galindo por sus enseñanzas y recomendaciones.

A mis amigos y compañeros con quienes compartí mi vida universitaria.

Adrián Laje quien ha sido mi compañero y amigo durante el inicio y fin de la carrera universitaria incluso prepo.

Daliana, Luis, Henry, Kevin y Anthony, compañeros y amigos con quienes he trabajado y estudiado en todo el transcurso de la carrera en ciencias de la computación.

A Johan que me dijo que la carrera de Sistemas se llama ahora Computación y terminé 6 años aquí.

A Jimmy que me enseñó que no es necesario entrar a clases, hacer proyectos o dar exámenes para aprobar una materia.

Y mis amigos Alejandra, Yadira, Johan, Rasu, Kevin, Boris, Lauro, Wilman, Jeremy, David, Josselyn, Toa, Elisabeth, Yajaira, Gabriel, David, Wilson, Ignacio y Cristian.

ÍNDICE DE CONTENIDO

[CERTIFICACIONES I](#_Toc85388319)

[DECLARACIÓN DE AUTORÍA III](#_Toc85388320)

[DEDICATORIA IV](#_Toc85388321)

[AGRADECIMIENTO V](#_Toc85388322)

[ÍNDICE DE CONTENIDO VI](#_Toc85388323)

[RESUMEN VII](#_Toc85388324)

[ABSTRACT VIII](#_Toc85388325)

[1 INTRODUCCIÓN 1](#_Toc85388326)

[1.1 Objetivo general 1](#_Toc85388327)

[1.2 Objetivos específicos 1](#_Toc85388328)

[1.3 Alcance 1](#_Toc85388329)

[1.4 Marco teórico 1](#_Toc85388330)

[2 METODOLOGÍA 2](#_Toc85388331)

[3 RESULTADOS, CONCLUSIONES Y RECOMENDACIONES 3](#_Toc85388332)

[3.1 Resultados 3](#_Toc85388333)

[3.2 Conclusiones 4](#_Toc85388334)

[3.3 Recomendaciones 5](#_Toc85388335)

[4 REFERENCIAS BIBLIOGRÁFICAS 6](#_Toc85388336)

[5 ANEXOS 7](#_Toc85388337)

[ANEXO I 8](#_Toc85388338)

RESUMEN

El proyecto consistió en el desarrollo de un sistema de minería de datos del portal web de la Secretaría Nacional de Cualificaciones y Capacitación Profesional (SETEC) para el análisis de la demanda de cursos y/o perfiles de profesionales cualificados a nivel nacional. El proyecto fue desarrollado usando metodología para construcción de sistemas de minería de datos CRISP-DM con las fases de entendimiento del negocio para el análisis de los módulos del portal web SETEC, el entendimiento de datos que conforman cada módulo del portal web, preparación de datos, extracción datos usando técnicas de raspado web, limpieza de datos, transformación de tipo de variables y colección de estos para la creación de un DataWareHouse. Posteriormente se realizó el análisis de la demanda en base a la definición de indicadores estáticos y dinámicos para la construcción de un modelo estadístico que finalmente se lo desplegó en la herramienta Microsoft PowerBI. Adicional a las seis fases de CRISP-DM se tuvo una fase de visualización de datos en la herramienta PowerBI a la cual se le realizó análisis de usabilidad en base a los principios de usabilidad de Nielsen.

**PALABRAS CLAVE:** Web Scraping, CRISP-DM, SETEC, Data Analysis, DataWareHouse, PowerBI, R, R Studio, SQL, RapidMiner, Python, Usability, ETL, EDA.

ABSTRACT

(Máximo 250 palabras)

**KEYWORDS:** word1, word2, …, word6.

1. DESCRIPCIÓN DEL COMPONENTE DESARROLLADO

El presente trabajo consiste en el desarrollo e implementación de un sistema de minería de datos de la plataforma web SETEC para el análisis de la demanda mediante indicadores estáticos y dinámicos utilizando la metodología de CRISP-DM.

Sugerencias: Para esta sección se sugiere un mínimo de 375 y un máximo de 750 palabras.

* 1. Objetivo general

Desarrollar un sistema de minería de datos para el análisis de la demanda de cursos y perfiles del portal web de la SETEC utilizando indicadores estáticos y dinámicos.

* 1. Objetivos específicos

Los objetivos específicos se detallan en base a la metodología CRISP-DM (Cross-Industry Standard Process for Data Mining) la cual es utilizada para la construcción de sistemas de minería de datos con fases típicas de proyecto, las tareas involucradas en cada fase y una explicación de las relaciones entre estas tareas [4]. Estas fases serán:

1. Entendimiento del negocio: Análisis y entendimiento de los módulos del portal web SETEC, diseño de objetivos
2. Entendimiento de datos: Análisis y entendimiento de los datos en los módulos del portal web SETEC
3. Preparación de datos: Extracción, selección, limpieza, transformación, construcción e integración de datos del portal web
4. Modelamiento: Selección de técnicas de modelización, generación de un diseño de pruebas, construcción y aceptación de modelos para el análisis de la demanda en base a los datos preparados.
5. Evaluación: Evaluación de los modelos desarrollados en base a los objetivos planteados y revisión de activades, decisiones de las anteriores fases y determinación de siguientes pasos a realizar.
6. Despliegue: Diseño de plan de despliegue de los modelos, monitorización y mantenimiento del sistema de minería de datos desarrollado

El sistema tendrá dos fases adicionales que son:

1. Visualización: visualización de los resultados a través de dashboards
2. Evaluación de usabilidad: análisis de la usabilidad del sistema de minería de datos desarrollado en base a los principios de Nielsen.
   1. Alcance

La metodología CRISP-DM se la utilizará para el diseño, construcción y evaluación de un sistema de minería de datos en base a 6 fases que se detallan a continuación para la creación del Sistema de minería de datos de SETEC.

1. Entendimiento del negocio:

1.1. Análisis y entendimiento de los módulos del portal web SETEC

1.2. Análisis y entendimiento del mercado laboral nacional

1.3. Diseño de objetivos de negocio e indicadores estáticos y dinámicos

1.4. Producción del plan de proyecto

2. Entendimiento de datos:

2.1. Análisis y entendimiento de los datos en los módulos del portal web SETEC

2.2. Extracción y colección de datos

2.3. Descripción de datos

2.4. Análisis exploratorio de datos

2.5. Verificación de la calidad de datos

3. Preparación de datos:

3.1. Selección de datos

3.2. Limpieza de datos

3.3. Transformación de datos

3.4. Formateo de datos

3.5. Construcción de nuevos datos

3.6. Integración de datos

3.7. Diseño del Datawarehouse

3.8. Almacenamiento de datos

4. Modelamiento:

4.1. Selección de técnicas de modelamiento

4.2. Generación de diseño de pruebas

4.3. Construcción de modelos

4.4. Evaluación de modelos

5. Evaluación:

5.1. Evaluación del resultado de los modelos desarrollados en base a los objetivos planteados

5.2. Carga de información predictiva en el Datawarehouse

5.3. Revisión de las actividades y decisiones tomadas en las anteriores fases

5.4. Determinación de los siguientes pasos

6. Despliegue:

6.1. Diseño de plan de despliegue de modelos

6.2. Plan de monitorización y mantenimiento

6.3. Despliegue de los modelos de en herramientas de visualización de datos

El sistema tendrá dos fases adicionales a las de CRISP-DM, que son:

7. Visualización:

7.1. Selección de gráficos para visualización de datos y resultados

7.2. Creación de gráficos estáticos y dinámicos

7.3. Creación de tableros de gráficos explicativos de los datos y resultados

8. Evaluación de usabilidad:

8.1. Evaluación de la usabilidad por medio de las métricas de usabilidad de Nielsen.

8.2. Evaluación de la usabilidad por medio de SUS.

* 1. Marco teórico

Inteligencia de negocios:

La inteligencia de negocios (Business Intelligence, BI) es el conjunto de técnicas, herramientas y sistemas que se utilizan para recopilar, integrar, analizar y presentar información de una empresa u organización con el fin de mejorar la toma de decisiones y el rendimiento de la organización. [1]

La inteligencia de negocios incluye un amplio espectro de actividades, desde la recopilación y almacenamiento de datos hasta la visualización y el análisis de información para obtener conocimiento y comprensión. También puede incluir la creación de informes y paneles de control para monitorear el rendimiento y la actividad de la empresa en tiempo real.

La inteligencia de negocios se utiliza en todos los aspectos de una empresa u organización, desde la planificación y el análisis financiero hasta la toma de decisiones de marketing y la gestión de operaciones. También puede ser útil para mejorar la eficiencia y la productividad, reducir costos y aumentar la satisfacción del cliente. [1]

CRISP-DM:

CRISP-DM es un modelo de proceso para el análisis y el desarrollo de proyectos de inteligencia de negocios y minería de datos. CRISP-DM es un acrónimo de Cross-Industry Standard Process for Data Mining. [1] Se ha convertido en un estándar de la industria y es ampliamente utilizado en todo el mundo para guiar el proceso de análisis de datos y la toma de decisiones basadas en datos.

El modelo CRISP-DM incluye seis fases:

1. Comprensión del negocio: se define el problema de negocio o la oportunidad de negocio que se quiere abordar y se establecen los objetivos del proyecto.
2. Recopilación de datos: se recopilan y seleccionan los datos relevantes para el proyecto.
3. Preparación de datos: se limpian y se preparan los datos para el análisis.
4. Modelado: se utilizan técnicas de análisis de datos para construir modelos que puedan ayudar a predecir resultados o entender tendencias.
5. Evaluación: se evalúan los resultados y se determina si se han alcanzado los objetivos del proyecto.
6. Implementación: se toman medidas en base a los resultados del proyecto y se implementan cambios en el negocio.

CRISP-DM es un modelo flexible y puede ser adaptado a diferentes proyectos y entornos. Es importante tener en cuenta que el modelo no es lineal y es posible que se tenga que volver a visitar ciertas fases varias veces a lo largo del proceso.

Ciencia de Datos:

La ciencia de datos es una disciplina que se centra en el análisis y la interpretación de grandes conjuntos de datos para obtener conocimientos y conclusiones valiosas. Esta disciplina combina técnicas de matemáticas, estadísticas y ciencias de la computación para explorar y analizar datos de una amplia variedad de fuentes, como bases de datos, archivos y metadatos.

En la ciencia de datos utiliza herramientas y técnicas de análisis de datos, como el aprendizaje automático y el procesamiento de lenguaje natural, para descubrir patrones y tendencias en los datos y hacer predicciones basadas en esa información. Esta disciplina se utiliza en una amplia variedad de campos, como la salud, la financiación, la publicidad y el gobierno, para tomar decisiones informadas y mejorar la eficiencia y el rendimiento.

Microsoft Power BI:

Power BI es una plataforma de inteligencia empresarial (BI) que se utiliza para conectar, modelar y visualizar datos y crear informes personalizados y gráficos con los KPI clave. Además, permite obtener respuestas a preguntas comerciales utilizando inteligencia artificial para brindar resultados rápidos y precisos. Es una herramienta escalable y fácil de usar para el autoservicio. [3]

RapidMiner:

Python:

Python es un lenguaje de programación desarrollado a principios de la década de 1990 por Guido van Rossum el cual es un lenguaje programación de alto nivel e interpretado lo que significa que no es necesario compilar el código antes de ejecutarlo. Este lenguaje es usado ampliamente en el desarrollo de aplicaciones de software, el análisis de datos y el aprendizaje automático, entre muchas otras debido a que cuenta con una gran cantidad de bibliotecas y frameworks disponibles que facilitan el desarrollo de aplicaciones.

Selenium

Por completar

Entornos de programación

#### Jupyter Lab

Entorno de desarrollo integrado para el lenguaje de programación Python y otros lenguajes de programación y de marcado. Es una aplicación de escritorio o web que se utiliza para escribir y ejecutar código, así como para crear documentos interactivos que combinan código, texto, imágenes y gráficos.

#### Visual Studio Code

Visual Studio Code (VS Code) es un editor de código fuente y un entorno de desarrollo integrado (IDE, por sus siglas en inglés) desarrollado por Microsoft. Es una aplicación de escritorio gratuita y de código abierto que se utiliza para escribir, depurar y ejecutar código en una amplia variedad de lenguajes de programación.

#### Google Colab

Es un servicio de Google que proporciona un entorno de desarrollo en la nube para escribir y ejecutar código en Python y capaz de conectarse con Google Drive para el almacenamiento de archivos. Colab es especialmente útil para el análisis de datos y el aprendizaje automático, ya que proporciona acceso a GPUs y TPUs de alta potencia para acelerar el procesamiento.

#### RStudio

R Studio es un entorno de desarrollo integrado (IDE, por sus siglas en inglés) para el lenguaje de programación R. Es una aplicación de escritorio gratuita y de código abierto que se utiliza para escribir, ejecutar y depurar código R.

Lenguaje de programación en R

Desarrollado por Ross Ihaka y Robert Gentleman en los años 90, R es un lenguaje de programación y un entorno de desarrollo para el análisis estadístico de datos, visualización de datos, el aprendizaje automático e investigación científica. R es un lenguaje de programación de alto nivel, interpretado y uso libre que es compatible con una amplia variedad de sistemas operativos y se puede utilizar en conjunción con otras herramientas de análisis de datos, como Excel y SQL.

SQL:

SQL (Structured Query Language) es un lenguaje de programación utilizado para interactuar con bases de datos relacionales. Con SQL, es posible crear, modificar y eliminar bases de datos, así como recuperar y manipular datos almacenados en ellas.

MySQL Workbech:

Es una herramienta que permite a arquitectos de bases de datos, desarrolladores y administradores de bases de datos realizar tareas visualmente. Ofrece funciones completas para modelar datos, desarrollar consultas SQL y administrar el servidor, incluyendo la configuración del servidor, la gestión de usuarios y la realización de copias de seguridad. [4]

DataWarehouse:

Un DataWareHouse o almacenamiento de datos consiste en un repositorio central de información que permite un mejor análisis en la toma de decisiones. Los datos que lo componen son de sistemas transaccionales, bases de datos no relacionales, bases de datos no relaciones, documentos, sitios web, entre otros. El almacenamiento de datos es fundamental para el análisis de los datos de una organización o mercado a través de la creación de informes, paneles y otras herramientas de análisis. [2]

KPIs:

KPI es el acrónimo de Key Performance Indicator, que se traduce al español como Indicador Clave de Desempeño. Un KPI es una métrica o medida utilizada para evaluar el rendimiento o el progreso de una empresa o un proyecto en relación con un objetivo establecido. Los KPIs se utilizan a menudo para medir el éxito de una estrategia o para tomar decisiones de negocio. Estos requieren una medición, un objetivo, una fuente y una frecuencia de medición. Estos pueden ser numéricos, de progreso y/o cambio

Graphical user interface

Description automatically generated

Graphical user interface, timeline

Description automatically generated

A screenshot of a computer

Description automatically generated with low confidence

Se debe escoger el criterio en base a los objetivos del negocio

Utilizando la metodología SMART

Objetivos SMART

Formatos

#### XLSX

#### CSV

#### PDF

#### TXT

1. METODOLOGÍA
   1. Entendimiento del negocio:

Esta fase se caracterizó por el análisis del portal web SETEC acerca de los cursos, perfiles y certificados de cualificación registrados de personas capacitadas o certificadas por operadores de capacitación, organismos evaluadores de la conformidad o capacitadores independientes. El planteamiento de los objetivos a desarrollar entorno a análisis de la demanda de cursos o perfiles utilizando KPIs estáticos y dinámicos. Y la producción del plan del proyecto a realizarse.

Análisis y entendimiento de los módulos del portal web SETEC

El portal web SETEC cuenta con 7 módulos con información acerca de cursos, perfiles, organizaciones, capacitadores y personas capacitadas en distintas áreas y especialidades profesionales a nivel nacional como se muestra en la FIGURA.

Imagen que contiene Texto

Descripción generada automáticamente

Este portal cuenta con la siguiente estructura de módulos y submódulos:

**Catálogo Nacional de Cualificaciones – CNC**

Este módulo cuenta con la información de todos los perfiles de cualificación a nivel nacional de Operadores de Capacitación y Evaluadores de la Conformidad, así como los perfiles inhabilitados actualmente. Este módulo se compone de los siguientes submódulos.

* CNC para Operadores de Capacitación - CNCOC
* CNC para Organismos Evaluadores de la Conformidad - CNCOEC
* CNC Perfiles Inhabilitados - CNCPI
* Reportes
* Descargas

**Operadores de Capacitación – OC**

Corresponde a la información de los Operadores de Capacitación cuyo estado puede ser habilitado, suspendido, finalizado su vigencia de calificación y cancelados. Los submódulos que conforman este módulo son:

* Operadores de Capacitación – OC
* OC - Suspendidos – OCS
* OC - Finalizaron su vigencia de Calificación – OCF
* OC - Cancelados – OCC
* Reportes
* Descargas

**Personas Capacitadas por OC – PCOC**

Es el módulo de personas quienes han sido capacitadas en cursos y/o perfiles por Operadores de Capacitación. Los submódulos que lo conforman son:

* Búsqueda Personas Capacitadas – PCOC
* Reportes
* Descargas

**Organismos Evaluadores de la Conformidad** – OEC

Módulo con la información de Organismos Evaluadores de la Conformidad encargados de emitir certificaciones de perfiles a nivel nacional. Cuenta con la información de OEC habilitados, suspendidos, finalizados su vigencia de reconocimiento y cancelados. Los submódulos que lo conforman son:

* Organismos Evaluadores de la Conformidad – OEC
* OEC - Suspendidos – OECS
* OEC - Finalizaron su vigencia de Reconocimiento – OECF
* OEC - Cancelados – OECC
* Reportes
* Descargas

**Personas Certificadas por OEC – PCOEC**

Módulo que cuenta con la información de las Personas Capacitadas por Organismos Evaluadores de la Conformidad. Los submódulos que lo conforman son:

* Búsqueda Personas Certificadas – PCOEC
* Reportes
* Descargas

**Capacitadores Independientes – CI**

Este módulo cuenta con la información de capacitores independientes quienes imparten cursos de capacitación a profesionales a nivel nacional. Los submódulos que lo conforma son:

* Búsqueda de Capacitadores Independientes – CI
* Reportes
* Descargas

**Personas Capacitadas por CI – PCCI**

Son las Personas Capacitadas por Capacitadores Independientes quienes se han certificado en los cursos que han tomado. Los submódulos que lo conforman son:

* Búsqueda de Personas Capacitadas – PCCI
* Reportes
* Descargas

Análisis y entendimiento del mercado laboral nacional

En Ecuador se registró a nivel nacional, en enero 2022, la tasa de empleo adecuado fue de 33,1%; para el área urbana de 41,4%; mientras que, en el área rural fue de 17,3%. [3] Esta tasa de empleo está conformada por profesionales quienes cuentan con certificaciones de cualificación de SETEC (Secretaria de Cualificaciones), el SECAP ( Servicio Ecuatoriano de Capacitación Profesional). Estas entidades validan las cualidades profesionales en distintos sectores laborales como son (Gestión Documental, Formación de Formadores, Maquillaje, Prevención de Riesgos Laborales, Apicultor, entre otros). [4]

La probabilidad de conseguir un empleo adecuado aumenta con estas certificaciones oficiales de parte de estas instituciones gubernamentales. Estas certificaciones están disponibles para personas mayores de 16 años ecuatorianos, que cumplan los prerrequisitos establecidos para cada perfil de otros sectores, así como en la normativa legal vigente para el servicio de certificación de personas por competencias laborales. [4]

Ministerio del Trabajo a través de la Subsecretaría de Cualificaciones Profesionales entregó las resoluciones de Calificación como Operadores de Capacitación (OCC) y el Reconocimiento como Organismo de la Conformidad (OEC) a más de 60 institutos y empresas que cumplieron con los requisitos establecidos para este efecto. [5]

Estos nuevos OEC y OCC se suman a los 181 Organismos Evaluadores de la Conformidad Reconocidos y 424 Operadores de Capacitación Calificados con los que cuenta actualmente el Sistema Nacional de Cualificaciones Profesionales. [5] De acuerdo con el P L AN NACIONAL DE EDUC A CIÓN Y FORMACIÓN TÉCNICA Y PROFESIONAL emitido por el ministerio de educación y ministerio del trabajo. [6]

Diseño de objetivos de negocio e indicadores estáticos y dinámicos

Se deberá crear un sistema de minería de datos enfocado al análisis de la demanda de cursos y/o perfiles profesionales cualificados por SETEC utilizando indicadores (KPIs) estáticos y dinámicos.

Se deberá tener datos en gran cantidad y calidad para poder procesarlos y analizarlos.

Se deberá mostrar a través de panales analíticos (Dashboards) los KPIs seleccionados para el análisis de la demanda por parte de los profesionales en distintos sectores laborales y áreas geográficas.

En base al análisis desarrollado en la unidad 2.1.1 los KPIs a desarrollar son de los Organismos de Capacitación (OC), Organismos Evaluadores de la Conformidad (OEC) y Capacitadores Independientes (CI). Estos KPIs son:

* **Organismos de Capacitación (OC):**
  + Número de personas capacitadas o certificadas por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, horas invertidas en cursos o perfiles por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, ganancias de los cursos o perfiles por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, cursos por Razón Social por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Volumen de personas capacitadas por Razón Social, Fecha, Cantón, Provincia y Fecha.
  + Porcentaje de asistencia de los convocados a los cursos a través de redes sociales por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
* **Organismos Evaluadores de la Conformidad (OEC):**
  + Número de personas certificadas por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, horas invertidas en perfiles por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, ganancias de los perfiles por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, cursos por Razón Social por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Volumen de personas certificadas por Razón Social, Fecha, Cantón, Provincia y Fecha.
  + Porcentaje de asistencia de los convocados a los cursos a través de redes sociales por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
* **Capacitadores independientes (CI):**
  + Número de personas capacitadas por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, horas invertidas en cursos por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, ganancias de los cursos por Razón Social, Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Total, cursos por Razón Social por Área, Familia, Especialidad, Sector, Modalidad, Cantón, Provincia y Fecha.
  + Volumen de capacitados por Razón Social, Fecha, Cantón, Provincia y Fecha.

Producción del plan de proyecto

Para el desarrollo de este proyecto se deberá seguir el siguiente plan de 16 semanas en donde se deberá cubrir y avanzar en cada fase de desarrollo de la metodología CRISP-DM y las dos fases extra de visualización y análisis de usabilidad:

|  |  |  |
| --- | --- | --- |
| **Semana referencial /**  **Etapas** | **Tareas específicas** | **Resultado esperado (si aplica)** |
| **1** | **Investigación de la literatura relacionada**   * **Metodología** * **Herramientas** * **Conceptos teóricos y técnicos** | * **Recopilación de fuentes investigación** * **Desarrollo del marco teórico** |
| **2** | **Entendimiento del negocio**   * **Análisis y entendimiento de los módulos del portal web SETEC** * **Análisis y entendimiento del mercado laboral nacional** * **Diseño de objetivos de negocio e indicadores estáticos y dinámicos** * **Producción del plan de proyecto** | * **Informe de análisis de los módulos que conforman el portal web SETEC** |
| **3** | **Entendimiento de los datos**   * **Análisis y entendimiento de los datos en los módulos del portal web SETEC** * **Extracción y colección de datos** * **Descripción de datos** * **Análisis exploratorio de datos** * **Verificación de la calidad de datos** | * **Informe de análisis de los datos que manejan los módulos del portal web SETEC** * **Datos recolectados iniciales** * **Informe de datos recolectados iniciales** |
| **4** | **Preparación de los datos**   * **Selección de datos** * **Limpieza de datos** * **Transformación de datos** * **Formateo de datos** * **Construcción de nuevos datos** * **Integración de datos** * **Diseño del Datawarehouse** * **Almacenamiento de datos** | * **Conjuntos de datos preparados almacenados** * **Datawarehouse** |
| **5** | **Modelamiento**   * **Selección de técnicas de modelamiento** * **Generación de diseño de pruebas** * **Construcción de modelos** * **Evaluación de modelos** | * **Modelos evaluados y aprobados** * **Informe de modelos seleccionados, evaluados y aprobados** |
| **6** | **Evaluación**   * **Evaluación del resultado de los modelos desarrollados en base a los objetivos planteados** * **Carga de información predictiva en el Datawarehouse** * **Revisión de las actividades y decisiones tomadas en las anteriores fases** * **Determinación de los siguientes pasos** | * **Información predictiva almacenada en el Datawarehouse** * **Informe de evaluación de modelos** |
| **7** | **Despliegue**   * **Diseño de plan de despliegue de modelos** * **Plan de monitorización y mantenimiento** * **Despliegue de los modelos de en herramientas de visualización de datos** | * **Modelos desplegados** * **Informe de despliegue de modelos** |
| **8** | **Visualización** | * **Gráficos de datos y resultados** |
| **9** | **Evaluación de usabilidad** | * **Informe de evaluación de métricas de Nielsen** * **Informe de evaluación de SUS** |
| **10** | **Documentación y correcciones** | * Trabajo de Integración Curricular |
| **11** | **Documentación y correcciones** | * Trabajo de Integración Curricular |
| **12** | **Documentación** | * Trabajo de Integración Curricular |

* 1. Entendimiento de datos:

Análisis y entendimiento de los datos en los módulos del portal web SETEC

Los datos que cuenta el portal web SETEC varían de acuerdo con cada módulo y submódulo que lo conforman. Los datos relevantes para su extracción y colección se encuentran en los siguientes módulos y submódulos:

1. Catálogo Nacional de Cualificaciones
   1. CNC para Operadores de Capacitación – CNCOC

Columnas de datos:



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |
| --- | --- | --- | --- |
| Perfil | Competencia Laboral | Familia | Sector |

* 1. CNC para Organismos Evaluadores de la Conformidad – CNCOEC

Columnas de datos:

Text

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |
| --- | --- | --- | --- |
| Perfil | Competencia Laboral | Familia | Sector |

* 1. CNC Perfiles Inhabilitados – CNCPI

Columnas de datos:



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| Perfil | Familia | Sector |

1. Operadores de Capacitación- OC
   1. Operadores de Capacitación- OC

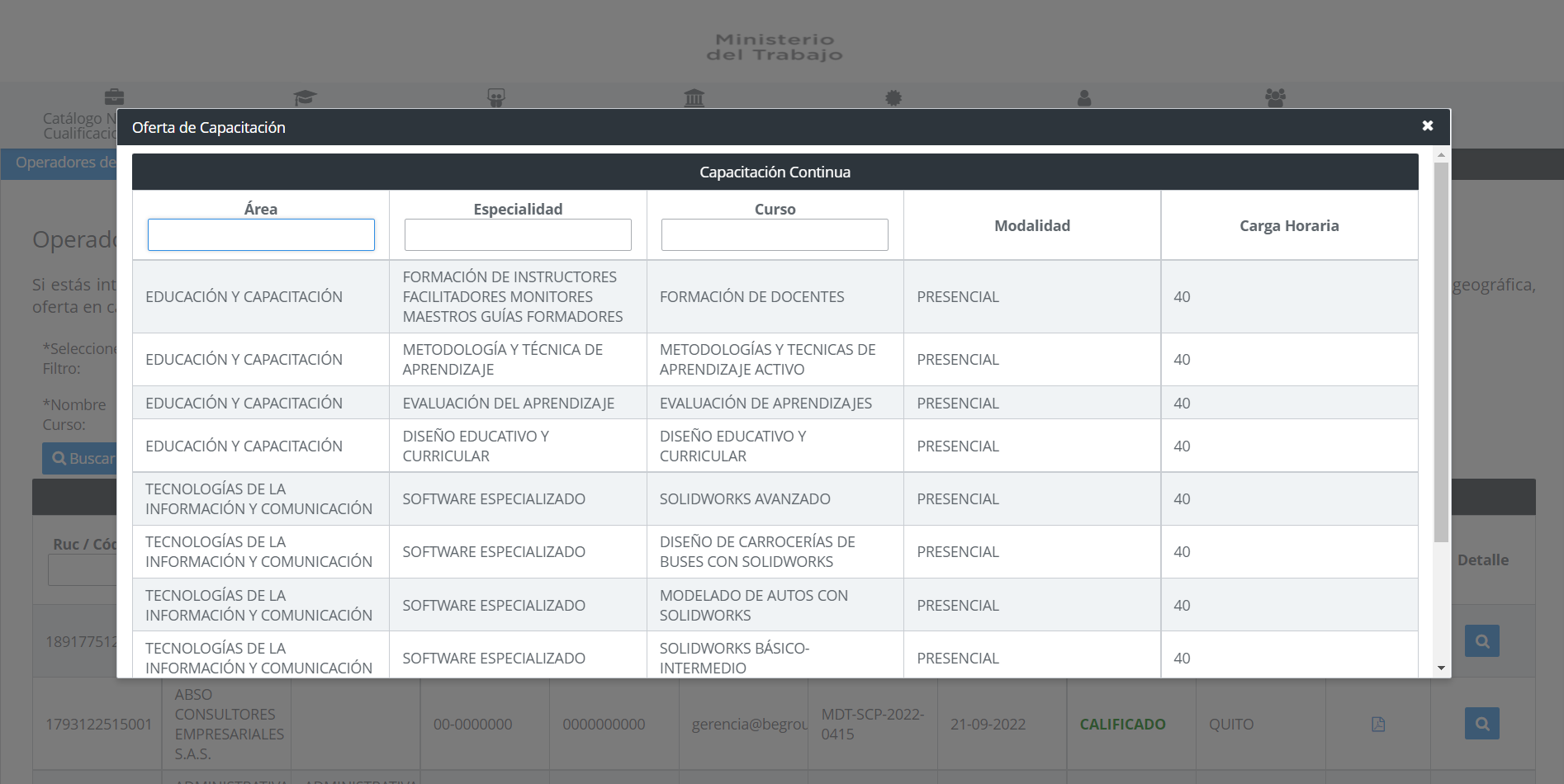
Columnas de datos:

Text

Description automatically generated

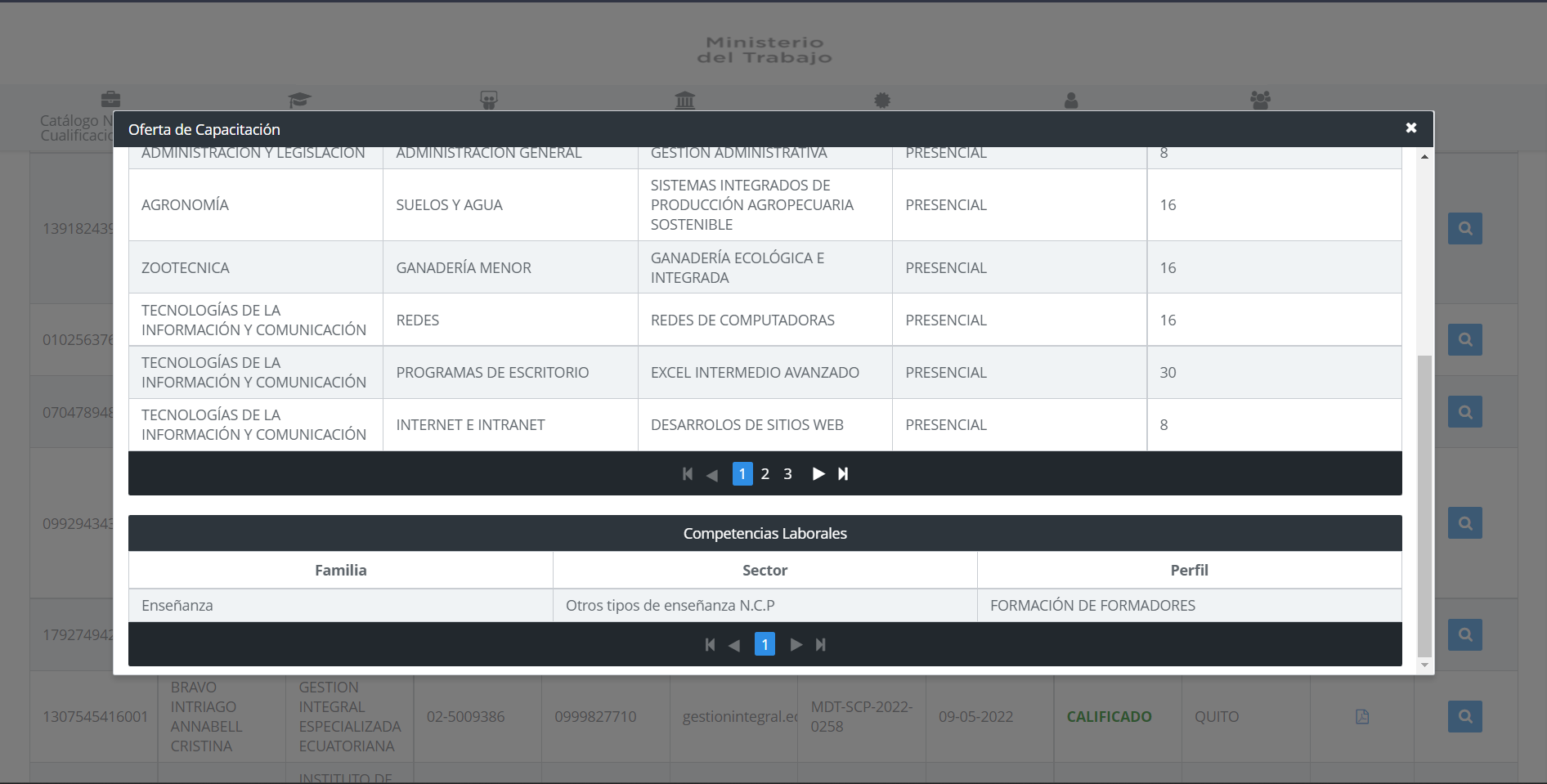
**Datos relevantes**

Al dar clic en **Ver Detalle** se obtiene los datos de **Capacitación Continúa** (OC\_CC):



**Datos relevantes**

Además de los datos de **Competencias Laborales (OC\_CL)** :



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| RUC / Código | Razón Social | Nombre Comercial | Teléfono | Celular | Email | Número Resolución | Fecha Resolución | Estado | Cantón | Detalle |

Detalle OC\_CC

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Área | Especialidad | Curso | Modalidad | Carga Horaria |

Detalle OC\_CL

|  |  |  |
| --- | --- | --- |
| Familia | Sector | Perfil |

* 1. OC - Suspendidos – OCS

Columnas de datos:



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OC - Finalizaron su vigencia de Calificación – OCF

Columnas de datos:



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OC - Cancelados – OCC

Columnas de datos:

Graphical user interface, text, application

Description automatically generated

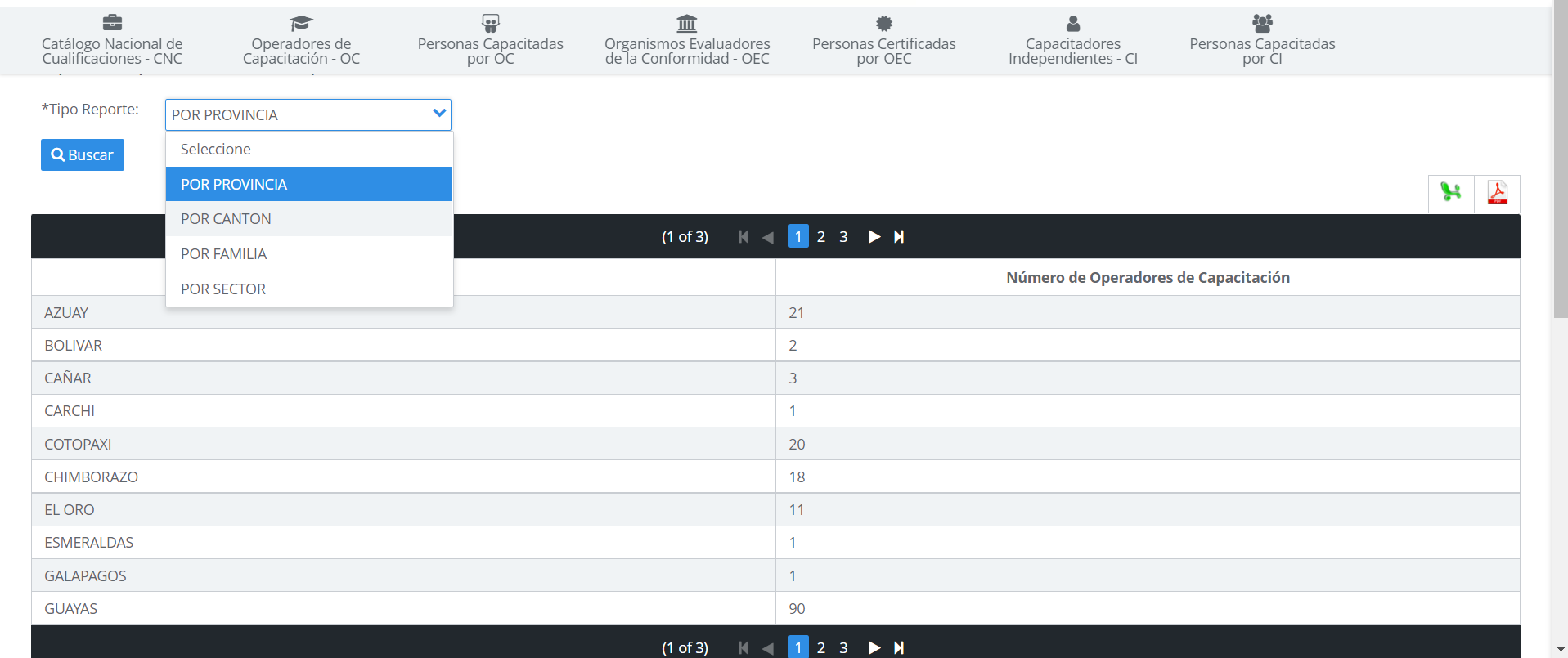
**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OC – Reportes

En este submódulo se pueden generar reportes de el número de Razones sociales por Provincia, Cantón, Familia o Sector. Estos reportes pueden ser descargados en formato .xlsx o .pdf



* 1. OC - Descargas

En este submódulo se puede descargar cierta información adicional de los OC como son:

Nombre de la Razón Social, su Provincia y Cantón

Graphical user interface

Description automatically generated with medium confidence

Nombre de la Razón Social, Áreas y Especializaciones que cuentan

Graphical user interface, text

Description automatically generated

Nombre de Razón Social, Familia, Sector y Perfil con los que cuentan

Graphical user interface, text, application

Description automatically generated

Estas tablas de datos se pueden descargar en formato .xlsx

1. Personas Capacitadas por OC – PCOC
   1. Búsqueda Personas Capacitadas

Columnas de datos:

Text

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Número Documento | Apellidos / Nombres | Tipo Capacitación | Nombre Curso / Perfil | Número Horas | Razón Social OC | Nombre Comercial OC |

* 1. Reportes

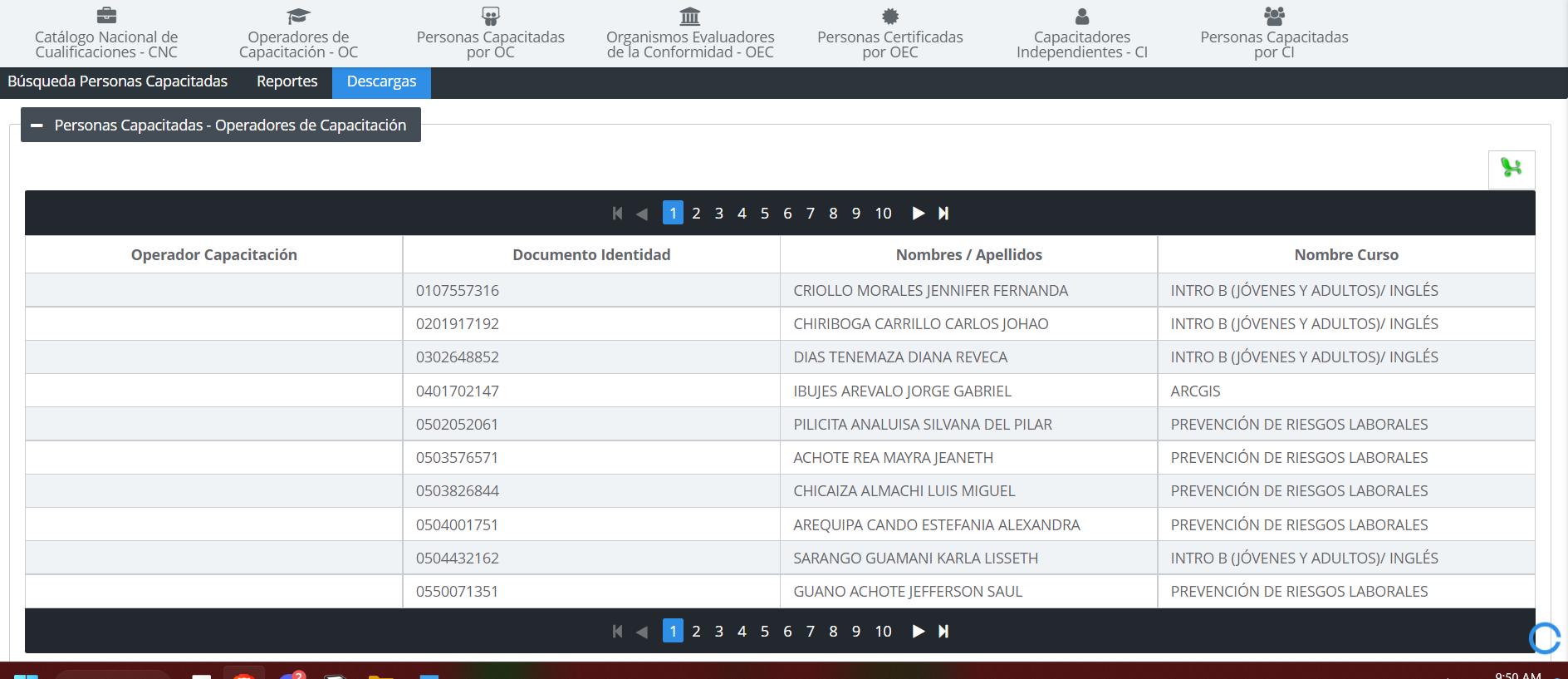
Este submódulo cuenta con reportes del número de personas capacitadas por Razón Social (Operador), Tipo de Capacitación y Nacionalidad. Estos reportes pueden ser descargados en formato .xlsx y .pdf

A picture containing table

Description automatically generated

* 1. Descargas

En submódulo cuenta con una tabla de datos de todos las personas capitadas hasta la fecha actual. Los datos que cuenta son: Operador Capacitación, Documento Identidad (de la persona capacitada), Nombres/Apellidos (de la persona capacitada) y Nombre del Curso (El que fueron capacitados). En teoría la pagina permite descargas estos datos en formato .xlsx



1. Organismos Evaluadores de la Conformidad – OEC
   1. Organismos Evaluadores de la Conformidad – OEC

Columnas de datos:

Text

Description automatically generated

**Datos relevantes**

Al dar clic en **Ver Detalle** se obtiene el **Detalle de Reconocimiento OEC (OEC\_DR)** con los siguientes datos:

Graphical user interface, text

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

**Submódulo principal**

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| RUC / Código | Razón Social | Dirección | Teléfono | Celular | Email | Número Resolución | Fecha Resolución | Cantón | Estado Resoluación |

**Detalle**

|  |  |
| --- | --- |
| Perfil | Esquema de Certificación |

* 1. OEC - Suspendidos – OECS

Columnas de datos:

Graphical user interface, text

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OEC - Finalizaron su vigencia de Reconocimiento – OECF

Columnas de datos:

Graphical user interface, text, application

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OEC - Cancelados – OECC

Columnas de datos:



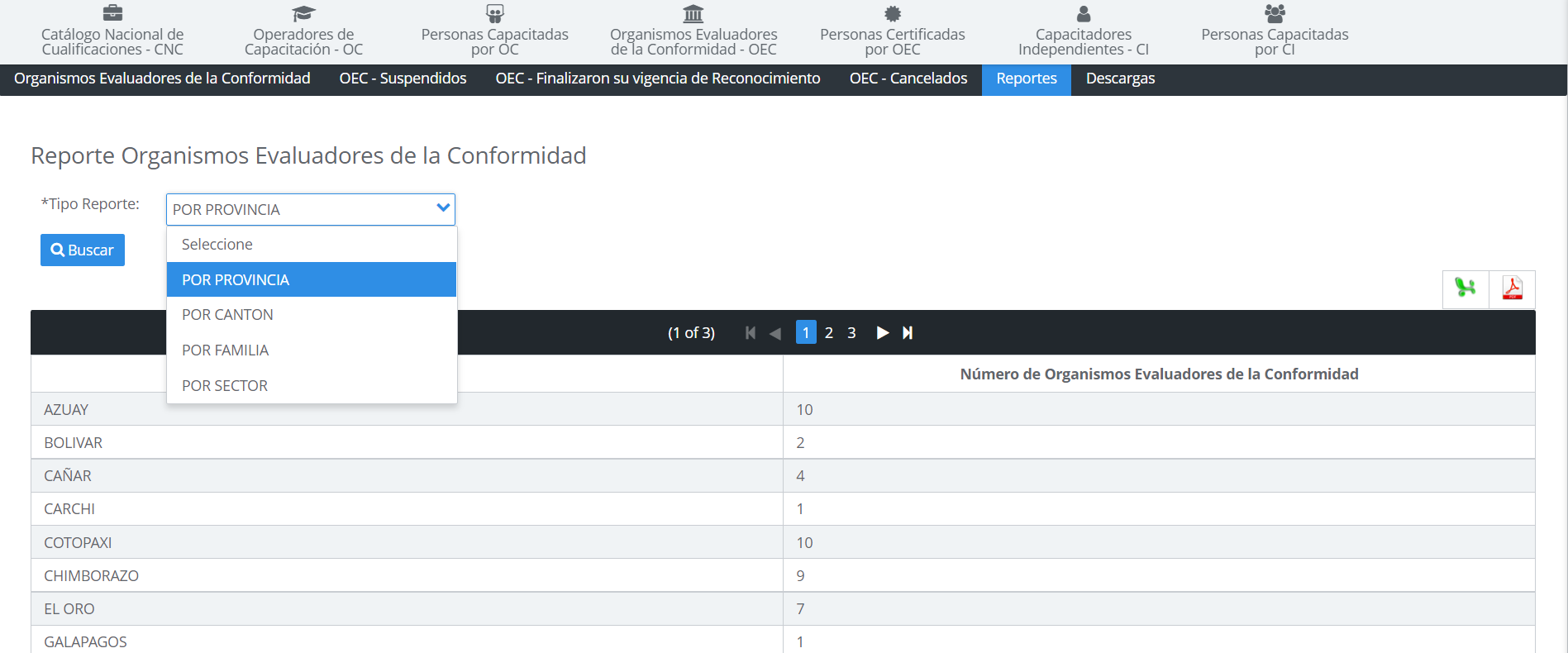
**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |
| --- | --- | --- |
| RUC / Código | Razón Social | Estado |

* 1. OEC – Reportes

Este submódulo cuenta con los reportes del número de Razones Sociales de Organismos Evaluadores de la Conformidad por Provincia, Cantón, Familia y Sector. Estos reportes pueden ser descargados en formato .xlsx y .pdf



* 1. OEC – Descargas

Este submódulo cuenta con la información a descargar de

Falta desarrollar

1. Personas Certificadas por OEC – PCOEC

Columnas de datos:

Graphical user interface, text, application

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Nombres | Perfil | Número Certificación | Fecha Certificación | Fecha Vigencia | Celular | Ruc Organismo Evaluador | Organismo Evaluador |

1. Capacitadores Independientes – CI

Columnas de datos:

Graphical user interface, text, application

Description automatically generated

**Datos relevantes**

Al dar clic en **Ver Detalle** se obtiene la **Lista de Cursos Aprobados (CI\_LCA)** con los siguientes datos:

Graphical user interface, application

Description automatically generated

**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Número Documento | Apellidos/Nombres | Código Resolucion | Vigencia Resolucion | Email | Celular |

Detalle

|  |  |  |  |
| --- | --- | --- | --- |
| Nombre Curso | Área | Especialidad | Carga Horaria |

1. Personas Capacitadas por CI – PCCI

Columnas de datos:



**Datos relevantes**

Datos relevantes por extraer:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Número Documento | Apellidos/Nombres | Nombre Curso | Número horas | Instructor/Capacitador |

Para desarrollar la estimación de precios de los cursos ofertados en el portal web se tomó como referencia los cursos ofertados por el Centro de Educación Continua (CEC – EPN). Se tomó únicamente los datos de cursos en el sector de TICs como se muestra en la siguiente figura.

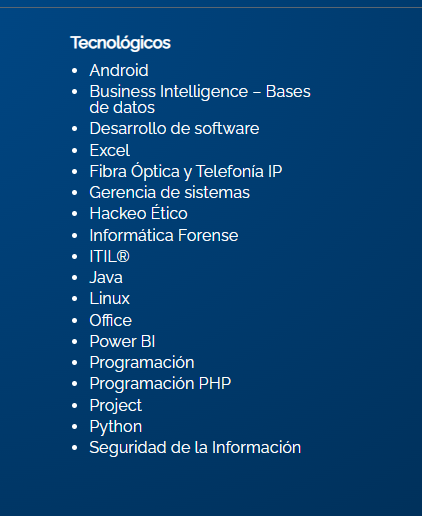


Fig. 1 – Lista de cursos tecnológicos ofertados por el CEC en 2022B [7]

Extracción y colección de datos

**Proceso de Extracción de datos**

**Análisis del Portal Web**

El proceso de extracción y colección de datos fue un proceso extenso y complicado debido a la naturaleza del portal web cuya conexión inestable fue la razón por la que:

* Está fuera de servicio muy frecuentemente.
* Falla al realizar consultas que son muy pesadas en cuanto a cantidad de datos.
* No permite una navegabilidad fluida
* No es posible descargar archivos cuyos volúmenes son muy extensos

Estos aspectos fueron tomados en cuenta al momento de crear los scripts para la extracción de datos, así como los archivos que es posible descargar del portal sin que este falle y presente este mensaje.

Graphical user interface, website

Description automatically generated

También el tiempo de sesión del sitio web por usuario es aproximadamente de 60 minutos, una vez trascurrido ese tiempo se mostrará el mensaje de sesión caducada.

Graphical user interface, website

Description automatically generated

La estructura de la página cuenta con:

* Módulos: Secciones del portal con la información delos CNC, OC, PCOC, OEC, PCOEC, CI y PCCI
* Submódulos: Secciones del módulo con la información respectiva sean datos generales, reportes o descargas. Estos pueden variar por módulo
* Selectores de búsqueda: Permiten seleccionar el tipo de búsqueda a realizar en el submódulo para posteriormente hacer la consulta de los datos
* Botones de búsqueda y cancelación: Respectivamente el primer botón permite ejecutar la consulta desarrollada y el botón de cancelación elimina la consulta anterior o la consulta seleccionada.
* Página: Consta de los datos a extraer como son la tabla de datos, la cabecera y los detalles que son desplegados del botón de detalle
* Cabecera: Cuenta con los títulos de la columnas de datos
* Datos: Consta con la tabla de datos a extraer
* Botón de detalle: Este despliega información de los cursos y/o perfiles con los que cuenta un razón social (Este botón está presente únicamente en CNC, OC, OEC y CI)
* Controladores de página: Estos permiten navegar entre las páginas del submódulo

Timeline

Description automatically generated

**Proceso de web scrapping**

El proceso de extracción se la desarrollo con la técnica de *Web Scraping*(Raspado Web) creando un “robot” o “araña” encargado de extraer los datos de cada módulo, submódulo y detalle. Estos robots se desarrollaron en el lenguaje ***Python*** con la librería ***Selenium***, mientras que los datos recolectados se los almaceno en archivos ***.CSV*** utilizando la librería ***Pandas***. Todos los scripts de ***Python*** para web scrapping están adaptados a cada uno de los submódulos.

De modo que el tipo de consultas que se hizo a cada página varió para obtener la mayor cantidad de datos como se muestra a continuación:

**Datos existentes**

Esto incluye una amplia variedad de datos como son los sitios web y los archivos xlsx a descargar.

Los datos obtenidos fueron recolectados desde el 4 de enero de 2023 hasta el 8 de enero de 2023. Debido a que los datos de la página son actualizados día tras día es posible que los datos varíen de la estimación inicial del número de datos a obtener.

Los archivos obtenidos de la recolección de datos fueron los siguientes:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Scripts** | **Módulo** | **Submódulo** | **Detalle** | **Nombre de archivo** | **Tamaño del archivo** |
| ws\_cnc\_oc.py | CNC | CNCOC | - | cncoc.csv | 227 kB |
| ws\_cnc\_oec.py | CNC | CNCOEC | - | cncoec.csv | 70 kB |
| ws\_cnc\_pi.py | CNC | CNCPI | - | cncpi.csv | 36 kB |
| ws\_oc.py | OC | OC | - | oc.csv | 39 kB |
| ws\_ocs.py | OC | OCS | - | ocs.csv | 92 Bytes |
| ws\_ocs.py | OC | OCF | - | ocf.csv | 15 kB |
| ws\_occ.py | OC | OCC | - | occ.csv | 1 kB |
| ws\_oc\_oc.py | OC | OC | OC\_CC | oc\_cc.csv | 445 kB |
| ws\_oc\_cl.py | OC | OC | OC\_CL | oc\_cl.csv | 18 kB |
| ws\_pcoc.py | PCOC | PCOC | - | pcoc.csv | 45.3 MB |
| ws\_oec.py | OEC | OEC | - | oec.csv | 47 kB |
| ws\_oecs.py | OEC | OECS | - | oecs.csv | 367 Bytes |
| ws\_oecf.py | OEC | OECF | - | oecf.csv | 9 kB |
| ws\_oecc.py | OEC | OECC | - | oecc.csv | 1 kB |
| ws\_oec\_dr.py | OEC | OEC | OEC\_DR | oec\_dr.csv | 150 kB |
| ws\_pcoec.py | PCOEC | PCOEC | - | pcoec.csv | 75 MB |
| ws\_ci.py | CI | CI | - | ci.csv | 421 kB |
| ws\_ci\_lca.py | CI | CI | CI\_LCA | ci\_lca.csv | 1.1 MB |
| ws\_pcci.py | PCCI | PCCI | - | pcci.csv | 17.1 MB |

Adicional a los datos extraídos a través de web scrapping se descargó los archivos de los submódulos de descargas para completar los datos toda la base de datos a ser analizada. Los archivos obtenidos fueron:

|  |  |  |
| --- | --- | --- |
| Módulo | Nombre del archivo | Tamaño de archivo |
| OC | oc\_dl\_area\_especialidad.csv | 293 kB |
| OC | Oc\_dl\_familia\_sector\_perfil.csv | 38 kB |
| OC | Oc\_dl\_provincia\_canton.csv | 31 kB |
| OEC | Oec\_dl\_familia\_sector\_perfil.csv | 378 kB |
| OEC | Oec\_dl\_provincia\_canton.csv | 15 kB |
| CI | Ci\_dl\_provincia\_canton.csv | 77 kB |

**Datos adicionales**

Para el cálculo de costos y horas de cursos se utilizó datos de referencia de archivos de costos de educación continua, así como se requirieron de datos extra para la ubicación geográfica. Algunos de estos archivos fueron descargados de sitios web oficiales, otros fueron producto de investigación de mercado por parte de empresas o por invesgación autoría propia.

|  |  |  |
| --- | --- | --- |
| Nombre del archivo | Descripción | Fuente |
| costos\_oferta\_educacion\_continua.xlsx | Investigación de mercado de empresa externa desarrollado en 2021 que recopila información acerca de la oferta de educación continua de distintas instituciones a nivel nacional | Documento de investigación |
| costos\_de\_cursos\_cec\_epn.xlsx | Investigación de información de cursos de centro de educación continúa de la Escuela Politécnica Nacional desarrollada en el año 2022 | Documento desarrollado por autoría propia.  URL: https://www.cec-epn.edu.ec/ |
| Anexos\_y\_Tablas\_para\_entrega\_Catastros\_GADS.xlsx | Documento con la información de las provincias, cantones y parroquias del Ecuador | Documento descargado del sitio oficial del SRI.  URL: https://www.sri.gob.ec/DocumentosAlfrescoPortlet/descargar/a7ce61fa-d8e6-4b77-8999-b99b617780a2/Anexos+y+Tablas+para+entrega+Catastros+GADS.xlsx |

Descripción de datos

**Cantidad de datos**

Los datos recolectados se estiman en base a la cantidad de registros por página.

Nota: la formula para la cantidad de datos estimada es de numero de registros por numero de paginas del submódulo. Dependiendo del submódulo los registros por página son de entre 10 o 20 registros, salvo la última página cuya cantidad de registros puede variar entre 1 a 10 o 20 registros. En cuanto a los datos que formaban parte del detalle su número no es posible de estimar en concreto, esto pasa con los módulos OC\_CC, OC\_CL, OEC\_DR y CI\_LCA

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| Módulos/ Submódulo | Registros | Páginas | Total registros | Archivos | Filas | Columnas | Total datos |
| CNCOC | 20 | 33 | 660 | cncoc.csv | 659 | 4 | 2636 |
| CNCOEC | 20 | 11 | 220 | cncoec.csv | 209 | 4 | 836 |
| CNCPI | 10 | 27 | 270 | cncpi.csv | 280 | 3 | 840 |
| OC | 20 | 28 | 560 | oc.csv | 542 | 10 | 5420 |
| OCS | 3 | 1 | 3 | ocs.csv | 1 | 3 | 3 |
| OCF | 10 | 26 | 260 | ocf.csv | 178 | 3 | 534 |
| OCC | 10 | 2 | 20 | occ.csv | 12 | 3 | 36 |
| OC\_CC | - | - | - | oc\_cc.csv | 5217 | 7 | 36519 |
| OC\_CL | - | - | - | oc\_cl.csv | 100 | 5 | 500 |
| PCOC | 20 | 23892 | 477840 | pcoc.csv | 251989 | 8 | 2015912 |
| OEC | 13 | 20 | 260 | oec.csv | 221 | 10 | 2210 |
| OECS | 2 | 1 | 2 | oecs.csv | 4 | 3 | 12 |
| OECF | 10 | 11 | 110 | oecf.csv | 100 | 3 | 300 |
| OECC | 10 | 2 | 20 | oecc.csv | 14 | 3 | 42 |
| OEC\_DR | - | - | - | oec\_dr.csv | 1205 | 4 | 4820 |
| PCOEC | 20 | 19320 | 386400 | pcoec.csv | 301572 | 8 | 2412576 |
| CI | 10 | 271 | 2710 | ci.csv | 2970 | 6 | 17820 |
| CI\_LCA | - | - | - | ci\_lca.csv | 7105 | 6 | 42630 |
| PCCI | 10 | 19150 | 191500 | pcci.csv | 146010 | 5 | 730050 |

En cuanto a su calidad de datos se tiene la siguiente información de cada archivo.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte cncoc.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| perfil | 659 | 0 | object | GESTIÓN EN FABRICACIÓN DE CALZADO |
| competenecia\_laboral | 659 | 0 | object | PLANIFICAR, ORGANIZAR Y CONTROLAR LA EJECUCIÓN DE FABRICACIÓN DE CALZADO BAJO NORMAS TÉCNICAS, ESTÁNDARES DE CALIDAD, MEDIO AMBIENTE Y SEGURIDAD Y SALUD EN EL TRABAJO, IDENTIFICANDO E IMPLEMENTANDO MEJORAS A LOS PROCESOS DE PRODUCCIÓN. |
| familia | 659 | 0 | object | PRODUCTOS TEXTILES, CUERO Y CALZADO |
| sector | 659 | 0 | object | FABRICACIÓN DE CUEROS Y PRODUCTOS CONEXOS |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte cncoec.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| perfil | 209 | 0 | object | MANTENIMIENTO ELECTROMECÁNICO DE VEHÍCULOS |
| competenecia\_laboral | 209 | 0 | object | REALIZAR OPERACIONES DE MANTENIMIENTO, MONTAJE DE ACCESORIOS Y MODIFICACIONES DEL VEHÍCULO EN EL ÁREA DE MECÁNICA, HIDRÁULICA, NEUMÁTICA Y ELECTRICIDAD ELECTRÓNICA AJUSTÁNDOSE A PROCEDIMIENTOS Y TIEMPOS ESTABLECIDOS, CONSIGUIENDO LA CALIDAD REQUERIDA Y EN CONDICIONES DE SEGURIDAD. |
| familia | 209 | 0 | object | VEHÍCULOS, AUTOMOTORES, CARROCERÍAS Y SUS PARTES |
| sector | 209 | 0 | object | COMERCIO Y REPARACIÓN DE VEHÍCULOS AUTOMOTORES Y MOTOCICLETAS |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte cncpi.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| perfil | 280 | 0 | object | OPERACIONES DE ANÁLISIS DE PROCESOS DE PRODUCCIÓN DE BEBIDAS |
| familia | 280 | 0 | object | PRODUCCIÓN INDUSTRIAL DE BEBIDAS Y TABACOS |
| sector | 280 | 0 | object | ELABORACIÓN DE BEBIDAS |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oc.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 542 | 0 | object | 102889730001 |
| razon\_social | 542 | 0 | object | MOGROVEJO NARVAEZ JHONI WILLIAM |
| nombre\_comercial | 455 | 87 | object |  |
| telefono | 542 | 0 | object | 00-0000000 |
| celular | 542 | 0 | object | 0 |
| correo\_electronico | 542 | 0 | object | centrodecapacitacionesaustro@gmail.com |
| numero\_resolucion | 542 | 0 | object | MDT-SCP-2021-0016 |
| fecha\_resolucion | 542 | 0 | object | 26-01-2021 |
| estado | 542 | 0 | object | CALIFICADO |
| canton | 542 | 0 | object | CUENCA |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte ocs.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 1 | 0 | int64 | 1711061364001 |
| razon\_social | 1 | 0 | object | VALENCIA MARTINEZ NELLY BEATRIZ |
| estado | 1 | 0 | object | SUSPENDIDO |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte ocf.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 178 | 0 | object | 1702716364001 |
| razon\_social | 178 | 0 | object | ESCOBAR GARCES MARCELA FELICIDAD |
| estado | 178 | 0 | object | FINALIZO VIGENCIA DE CALIFICACION |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte occ.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 12 | 0 | int64 | 1713782132001 |
| razon\_social | 12 | 0 | object | CAIZA LOGACHO EDISON FERNANDO |
| estado | 12 | 0 | object | CANCELADO DEFINITIVAMENTE |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oc\_cc.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| documento | 5217 | 0 | object | 102889730001 |
| razon\_social | 5217 | 0 | object | MOGROVEJO NARVAEZ JHONI WILLIAM |
| area | 5217 | 0 | object | MECÁNICA AUTOMOTRIZ |
| especialidad | 5217 | 0 | object | MECÁNICA GENERAL |
| curso | 5217 | 0 | object | MECÁNICA AUTOMOTRIZ |
| modalidad | 5217 | 0 | object | PRESENCIAL |
| carga\_horaria | 5217 | 0 | int64 | 120 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oc\_cl.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 100 | 0 | object | 1709260000000 |
| razon\_social | 100 | 0 | object | CHACHAPOYA CACHIGUANGO FREDDY BAYARDO |
| familia | 100 | 0 | object | Industrias manufactureras |
| sector | 100 | 0 | object | Elaboración de comidas y platos preparados |
| perfil | 100 | 0 | object | PREPARACIÓN GASTRONÓMICA DE COCINA |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oec.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 221 | 0 | object | 1801742642001 |
| razon\_social | 221 | 0 | object | NAVAS ESCOBAR FAUSTO ENRIQUE |
| direccion | 221 | 0 | object | LAS LIMAS S/N Y LOS HIGOS |
| telefono | 221 | 0 | object | 03-2855442 |
| celular | 221 | 0 | int64 | 9.84E+08 |
| correo\_electronico | 221 | 0 | object | faustoenriquenavas@hotmail.com |
| numero\_resolucion | 221 | 0 | object | MDT-SCP-2021-0073 |
| fecha\_resolucion | 221 | 0 | object | 3/4/2021 |
| canton | 221 | 0 | object | AMBATO |
| estado | 221 | 0 | object | HABILITADO |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oecs.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 4 | 0 | int64 | 1792257077001 |
| razon\_social | 4 | 0 | object | LIDERAZGO CAPACITACION Y CONSULTORIA DAVSOC COMPAÑIA LIMITADA |
| estado | 4 | 0 | object | SUSPENDIDO |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oecf.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 100 | 0 | object | 1091747029001 |
| razon\_social | 100 | 0 | object | FUNDEHS CONSULTING |
| estado | 100 | 0 | object | FINALIZO VIGENCIA DE RECONOCIMIENTO |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte oec\_dr.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| ruc\_o\_codigo | 14 | 0 | int64 | 1704980158001 |
| razon\_social | 14 | 0 | object | MIER MEJIA MARY JEANNETH |
| estado | 14 | 0 | object | CANCELADO DEFINITIVAMENTE |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte ci.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| numero\_documento | 2970 | 0 | int64 | 914804364 |
| apellidos\_nombres | 2970 | 0 | object | VALENZUELA MENDIETA RICHARD OSWALDO |
| codigo\_resolucion | 2970 | 0 | object | RESOLUCIÓN No. MDT-CI-CAL-2021-0115 |
| vigencia\_resolucion | 2970 | 0 | object | 3/8/2023 |
| correo\_electronico | 2970 | 0 | object | rickovm@gmail.com |
| celular | 2970 | 0 | object | 996521955 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Reporte ci\_lca.csv** | | | | |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| numero\_documento | 7105 | 0 | int64 | 914804364 |
| apellidos\_nombres | 7105 | 0 | object | VALENZUELA MENDIETA RICHARD OSWALDO |
| curso | 7034 | 71 | object | PREVENCIÓN DE RIESGOS LABORALES |
| area | 7034 | 71 | object | PROCESOS INDUSTRIALES |
| especialidad | 7034 | 71 | object | SEGURIDAD PREVENCIÓN DE RIESGOS E HIGIENE INDUSTRIAL |
| carga\_horaria | 7034 | 71 | float64 | 16 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Nombre de columna** | **# de datos NO nulos** | **# de datos nulos** | **Tipo de datos** | **Ejemplo** |
| numero\_documento | 146010 | 0 | object | 1713185427 |
| apellidos\_nombres | 146010 | 0 | object | LUIS EDGAR TIMBILA SALAS |
| curso | 146010 | 0 | object | CONTRATACIÓN PÚBLICA DE BIENES Y SERVICIOS, PARA PROVEEDORES DEL ESTADO |
| numero\_horas | 146010 | 0 | int64 | 32 |
| capacitador\_independiente | 146010 | 0 | object | CHAVEZ JARAMILLO JONATHAN AUGUSTO |

Análisis exploratorio de datos

Verificación de la calidad de datos

* 1. Preparación de datos:

Selección de datos

Limpieza de datos

Transformación de datos

Formateo de datos

Construcción de nuevos datos

Integración de datos

Diseño del Datawarehouse

Creación de modelos entidad relación

Almacenamiento de datos

* 1. Modelamiento:

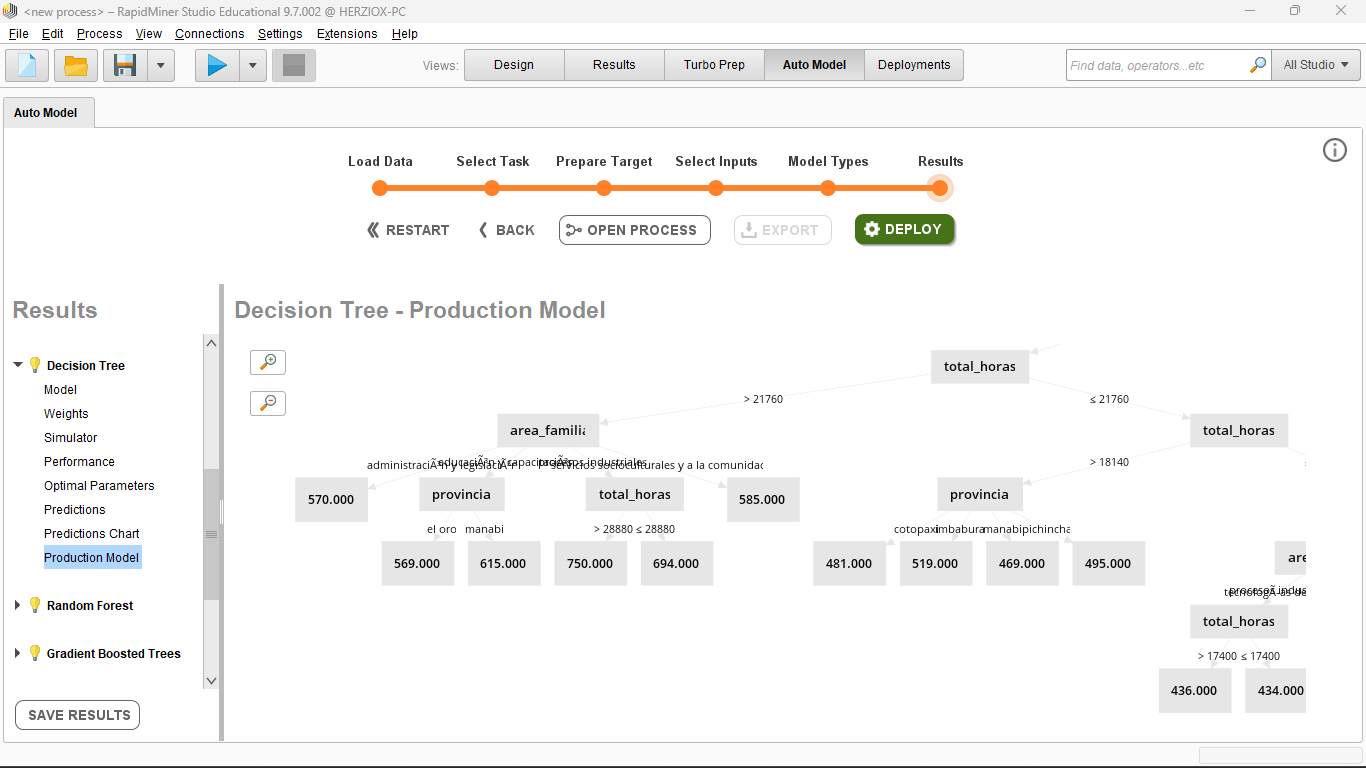
Selección de técnicas de modelamiento

Generación de diseño de pruebas

Construcción de modelos

Evaluación de modelos

Graphical user interface

Description automatically generated with low confidence

**Graphical user interface, application, table

Description automatically generated**

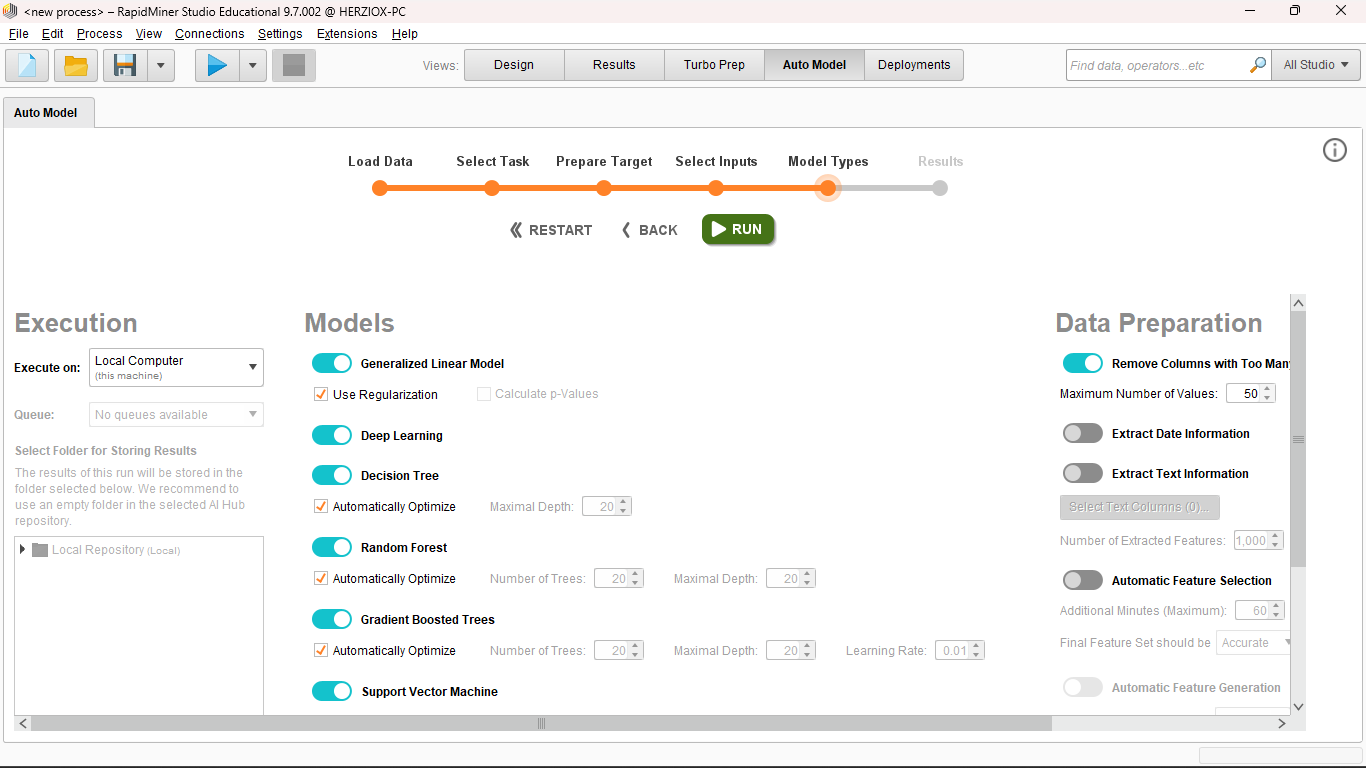
**Graphical user interface

Description automatically generated**

**Graphical user interface, text

Description automatically generated with medium confidence**

**Tipos de modelos**

****

**Graphical user interface, text, application, email

Description automatically generated**

**Selección de variables**

**Graphical user interface, text

Description automatically generated with medium confidence**

**Preparación de objetivos**

**Graphical user interface, text, application, email

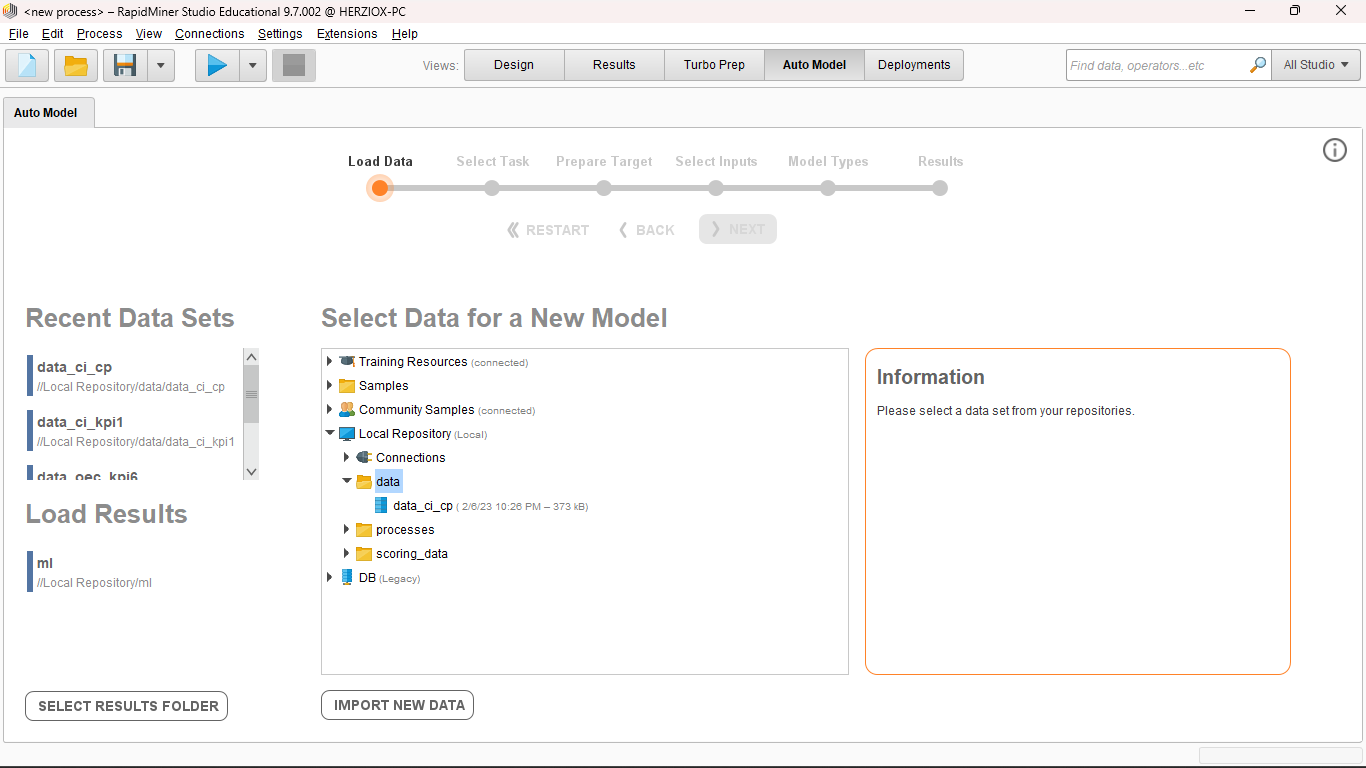
Description automatically generated**

**Selección de tareas**

**Graphical user interface

Description automatically generated**

**Cargar datos**

****

* 1. Evaluación:

Evaluación del resultado de los modelos desarrollados en base a los objetivos planteados

Carga de información predictiva en el Datawarehouse

Revisión de las actividades y decisiones tomadas en las anteriores fases

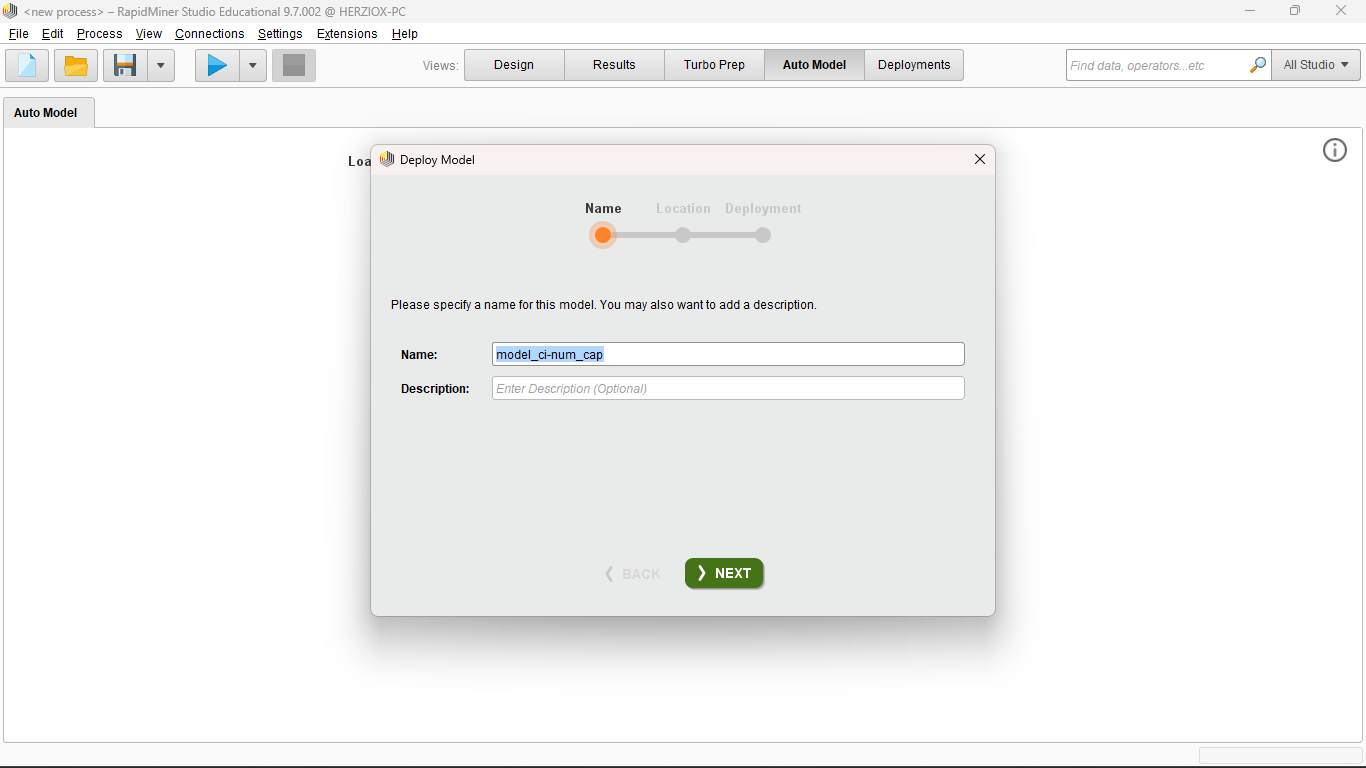
Determinación de los siguientes pasos

* 1. Despliegue:

Diseño de plan de despliegue de modelos

Plan de monitorización y mantenimiento

Despliegue de los modelos de en herramientas de visualización de datos





Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Graphical user interface, text, application

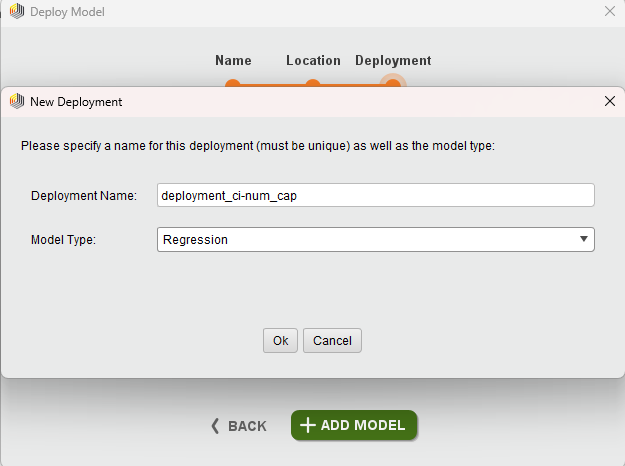
Description automatically generated

Graphical user interface, text, application

Description automatically generated

Graphical user interface, text, application

Description automatically generated



Graphical user interface, text, email

Description automatically generated

Graphical user interface, text, application, email

Description automatically generated

Table

Description automatically generated

Graphical user interface, table

Description automatically generated

* 1. Visualización:

Selección de gráficos para visualización de datos y resultados

Creación de gráficos estáticos y dinámicos

Creación de tableros de gráficos explicativos de los datos y resultados

* 1. Evaluación de usabilidad:

Evaluación de la usabilidad por medio de las métricas de usabilidad de Nielsen.

Evaluación de la usabilidad por medio de SUS.

1. PRUEBAS, RESULTADOS, CONCLUSIONES Y RECOMENDACIONES
   1. Pruebas

Establecer las pruebas realizadas que permiten determinar la validez del componente.

* 1. Resultados

Exponer los resultados obtenidos, utilizando para esto el apoyo de tablas, figuras, entre otros.

Un ejemplo de una tabla se presenta a continuación:

**Tabla 1.** Resultados de las pruebas realizadas

|  |  |  |
| --- | --- | --- |
| **No. Prueba** | **Resultado** | **Tiempo [s]** |
| 1 | 10 | 0.9 |
| 2 | 5 | 0.5 |

Un ejemplo de una figura se presenta a continuación:

**Figura 1.** Resultados de las pruebas realizadas

Un ejemplo de una ecuación se presenta a continuación:

**Ecuación 1.** Trinomio Cuadrado Perfecto

* 1. Conclusiones

Presenta lo novedoso del trabajo de integración curricular, así como evaluación del cumplimiento o no de lo propuesto en los objetivos. En el caso en que no se cumpla uno o varios objetivos, y no se logren los resultados esperados, se propone una posible respuesta que explique por qué sucedió esto o las falencias de la planteado.

* 1. Recomendaciones

Indicar las recomendaciones formuladas a partir del desarrollo de este trabajo de integración curricular.

.

1. REFERENCIAS BIBLIOGRÁFICAS

Listar todas las referencias bibliográficas citadas en el documento, siguiendo las normas de estilo IEEE.

|  |  |
| --- | --- |
| [1] | IBM, «CRISP-DM Help Overview,» BM, 17 08 2021. [En línea]. Available: https://www.ibm.com/docs/en/spss-modeler/saas?topic=dm-crisp-help-overview. [Último acceso: 25 12 2022]. |
| [2] | Amazon Web Services, «Conceptos relacionados con el almacenamiento de datos,» Amazon Web Services, 2022. [En línea]. Available: https://aws.amazon.com/es/data-warehouse/. [Último acceso: 06 10 2022]. |
| [3] | INEC, «Encuesta Nacional de Empleo, Desempleo y Subempleo (ENEMDU), enero 2022,» INEC, Quito, 2022. |
| [4] | SECAP, «Certificación de personas por competencias laborales en otros sectores,» SECAP, 2022. [En línea]. Available: https://www.gob.ec/secap/tramites/certificacion-personas-competencias-laborales-otros-sectores. [Último acceso: 07 10 2022]. |
| [5] | ecuadorenvivo, «Ministerio de Trabajo calificó a nuevos Operadores de Capacitación,» ecuadorenvivo, 23 09 2022. [En línea]. Available: https://www.ecuadorenvivo.com/index.php/economia/item/150135-ministerio-de-trabajo-califico-a-nuevos-operadores-de-capacitacion. [Último acceso: 07 10 2022]. |
| [6] | Educación, Ministerio de; Trabajo, Ministerio de, Plan Nacional de Educación y Formación Técnica y Profesional, Quito: Ministerio de Educación, 2021. |
| [7] | CEC EPN, «Centro de Educación Continua EPN,» CEC EPN, 2022. [En línea]. Available: https://www.cec-epn.edu.ec/. [Último acceso: 15 11 2022]. |
| [8] | L. Carvajal, Metodología de la Investgación Científica. Curso general y aplicado, 28 ed., Santiago de Cali: U.S.C., 2006, p. 139. |

1. ANEXOS

En caso necesario, el documento escrito deberá incluir los anexos y secciones que incorporan información que sea relevante, pero que, por su extensión, no pueden ser incorporadas directamente en ninguna de las secciones anteriores. Normalmente, en la sección de Anexos se incluyen conjuntos de datos extensos, formatos de encuestas, entrevistas, enlaces hacia videos o programas que sean producto o formen parte del Trabajo de Integración Curricular, entre otros.

Ejemplo de Anexos se muestran a continuación:

ANEXO I. Conjunto de Datos Extensos

ANEXO II. Formato de Entrevista

ANEXO III. Enlaces

La numeración de los Anexos debe realizarse con números en formato romano.

ANEXO I

Incluir el contenido del Anexo I.

ASPECTOS DE FORMATO

Tipografía

Un resumen de la tipografía se presenta a continuación:

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Texto** | **Tamaño** | **Fuente** | **Estilo** | **Ejemplo** |
| Normal | 11 | Arial | Normal; Alineado justificado. | Texto de párrafo |
| Nivel 1 | 14 | Arial | Negrita; Enumerada; Mayúscula, Con sangría; Alineado izquierda. | **1 INTRODUCCIÓN** |
| Nivel 2 | 14 | Arial | Negrita; Enumerada; Con sangría; Alineado izquierda. | **1.2 Objetivo general** |
| Nivel 3 | 12 | Arial | Negrita; Alineado izquierda. | **Subsección** |

Espaciado

Todo el documento debe tener espaciado de 1,5. Las tablas pueden usar espaciado simple.

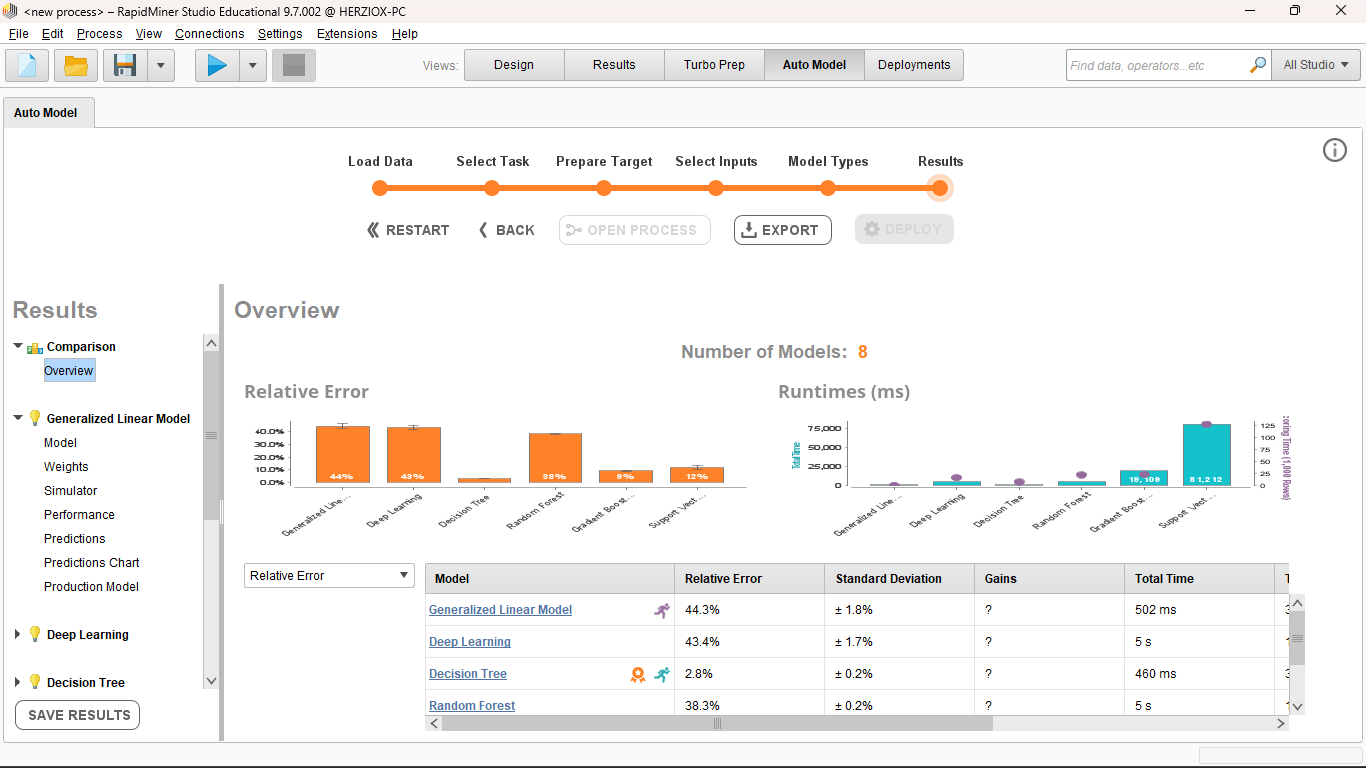
Numeración

Un resumen de la numeración se presenta a continuación:

|  |  |  |
| --- | --- | --- |
| **Elemento** | **Estilo** | **Número** |
| Tabla | Superior al elemento, Alineado centro. | Continuo, Número arábigo. |
| Figura | Inferior al elemento, Alineado centro | Continuo, Número arábigo. |
| Ecuación | Inferior al elemento, Alineado centro. | Continuo, Número arábigo. |
| Páginas | Inferior  Centro | Continuo, Número romano hasta Abstract y en arábigo hasta el final. |
| Anexos |  | Continuo, Número romano. |

* 1. Anexos

CI num\_cap



CI total\_horas

Graphical user interface

Description automatically generated with medium confidence

CI Total\_ganancias

Graphical user interface

Description automatically generated with medium confidence