Digital, Data Analytics, and Automation: Value Creation Through Digital E&P

19-20 NOVEMEBER 2024 | BANGKOK, THAILAND

Enhancing Carbon Capture & Storage Data Management: A Novel Integration of AI, Digital Technologies, and Cross-Industry Standard Process for Data Mining Framework

ADAM DANIEL BIN EFFENDI

SME & PE DATA ANALYST,
AEM ENERSOL SDN BHD

&

ZAYD UMAIR ZAKARIA

PE DATA ANALYST,
UPSTREAM TECHNOLOGY, DIGITAL & INNOVATION (UTDI), PETRONAS

Introduction and Project Objectives

1. Introduction to CCS Data Management:

- **Carbon Capture & Storage (CCS)** generates large volumes of data, but managing the associated data is complex.
- The project focuses on **centralizing Carbon Capture and Storage (CCS) data within the Storage Development Plan (SDP).**

2. Challenges in CCS Data Management:

- **Unstructured Data:** Logs and reports are often in diverse formats, making them difficult to access, standardize, and analyze effectively.
- **Governance, Security & Scalability:** Ensuring data integrity, compliance, secure access, and scalability is crucial as CCS projects grow. Efficient systems are needed to manage increasing data volumes without compromising security or performance.

3. Project Objectives:

- **Centralize CCS Data:** Create a single source of truth for seamless data access and improved consistency.
- **Automate with AI:** Leverage low code AI to streamline data extraction and processing, reducing manual effort.
- **Governance, Security & Insights**: Ensure data compliance, scalability, and security, while enabling quick, informed decision-making using structured data.

Main Component

CRoss Industry Standard Process-Data Mining (CRISP-DM):
Provides a structured, repeatable approach for data mining, ensuring consistency from data understanding to deployment. It defines each step, coordinating LLM and the low code AI within the workflow.

LLM (Large Language Model):
Helps process unstructured CCS data by generating REGEX patterns, reducing manual effort during data preparation. LLM outputs are used in the low code AI to automate data extraction based on recognized patterns.

Claude 3.5 Sonnet

Low Code AI & Xtralisis:

1. Automates Extraction, Transformation, and Loading (ETL) processes using Xtralisis workflow (Extract files -> Detect issues ; corrupt, duplicates -> Extract/analyze text -> Classify documents -> Deliver clean, structured data for quick analysis), streamlining data extraction and ensuring efficient data readiness for analysis.
2. Low Code AI application uses the pages/data identified by LLM.

Value Creation from AI and CRISP-DM Integration

Automation Efficiency

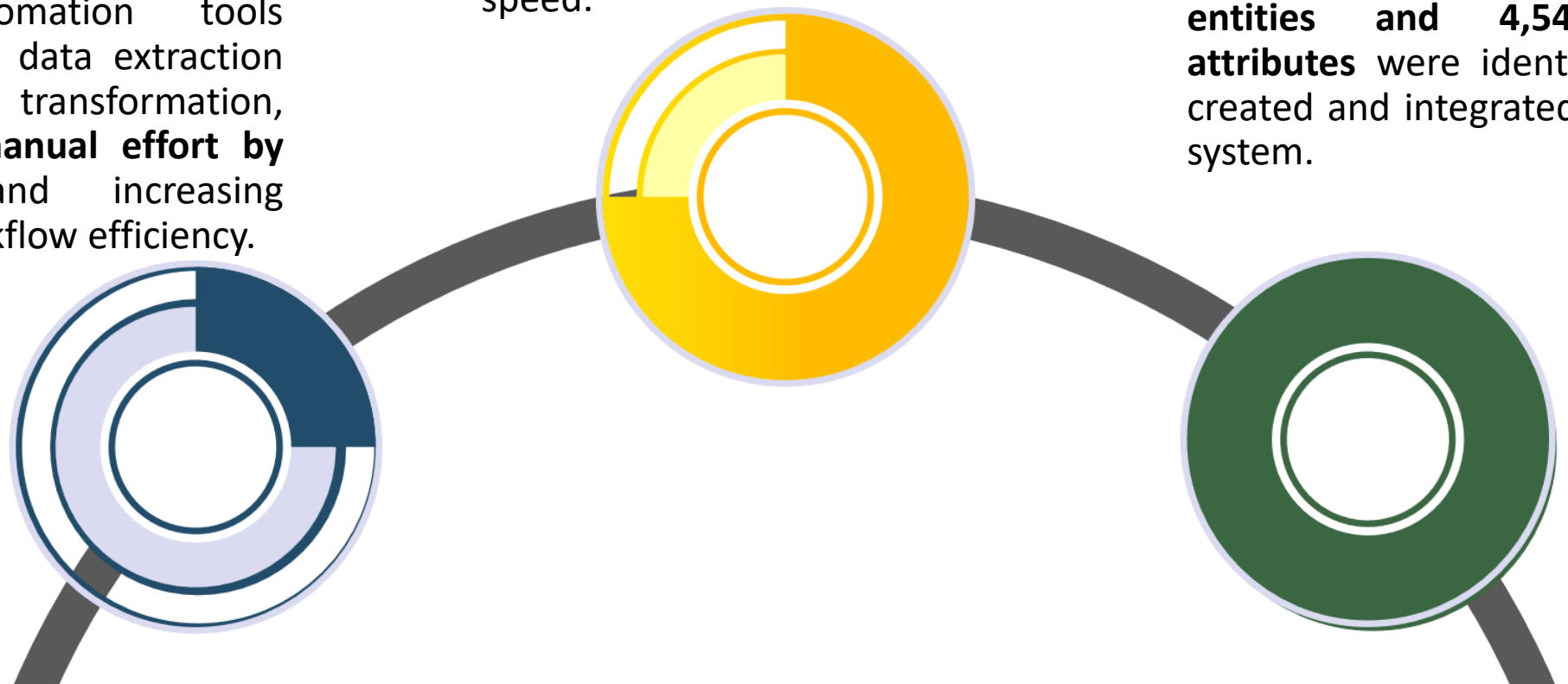
Integration of CRISP-DM with automation tools streamlined data extraction and transformation, **reducing manual effort by 10-30%** and increasing overall workflow efficiency.

Data Accessibility

Data can now be **accessed 10x faster**, allowing project teams to retrieve structured information right away and improve decision-making speed.

Data Recovery

Approximately **261 new data entities and 4,540 new attributes** were identified and created and integrated into the system.



AI & CRISP-DM Integration

Cross Industry Standard Process for Data Mining (CRISP-DM)

1. Business Understanding:

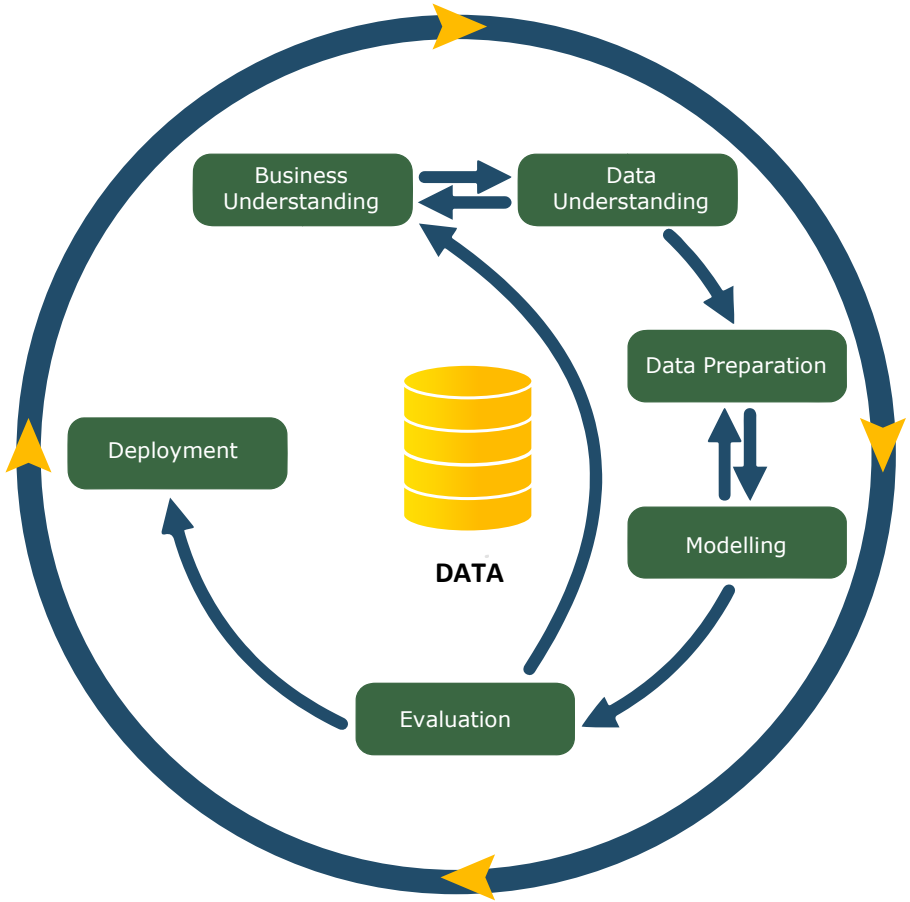
- **Project Goal**: Modernize CCS data management.
- **Alignment**: Focus on creating a **centralized database.**

2. Data Understanding:

- **Data Challenges**: The project faced **unstructured data**.
- **Assessment**: Identified key data points needed for analysis.

3. Data Preparation:

- **Automation**: Low code AI was used to **automate data extraction** and **transformation.**
- **Transformation**: Using **LLM-assisted REGEX**



4. Modeling:

- **Reduction of Manual Efforts**: Automation through low code AI.

5. Evaluation :

- **Model Evaluation**: Ensured that the automated data extraction met the quality.
- **Assessment**: Identified key data points needed for analysis.

6. Deployment:

- **System Implementation**: The processed and structured data was ingested into **CCS Storage Development Plan (SDP) database**.

AI & CRISP-DM Integration

Large Language Model (LLM)

1. **LLM for REGEX Creation**: LLMs helped generate and validate complex REGEX for more flexible and precise data extraction.
2. **Key Term Identification**: LLM helped create REGEX tailored to our understanding of the document, allowing efficient data location within large reports
3. **Data Security**: A locally deployed LLM ensured data privacy, reducing reliance on cloud-based services.

ATTRIBUTE	EXAMPLE VALUE	SONNET 3.5 CLAUD REGEX	GPT-4 REGEX	LLM INTERPRETATION DIFFERENCE
Well Name	ABC-1	^[A-Z0-9-]+$	([A-Z]+\-\d+)	SONNET: Fixed pattern. GPT-4: Capturing groups for naming conventions.
Bottom Logged Interval	1500.29	^\d+\.\d{2}$	(\d+\.\d{2,})	SONNET: Restricts to two decimals. GPT-4: Flexible with decimal places.
Log Description	MUD LOG TVD-MSL	^[A-Z\s-]+$	([A-Z\s]+LOG\s[TVD-MSL]+)	SONNET: Simple pattern. GPT-4: Context-aware for expected structures.
Scanned QC Remarks	ABC-1-1_MUDLOG 44.90-1500.29 TVD-MSL	^[A-Z0-9_\s.-]+$	([A-Z0-9_\-]+_MUDLOG\s\d+\.\d+-\d+\.\d+\sTVD\-MSL)	SONNET: Basic remarks. GPT-4: Captures complex fields, numeric ranges, and abbreviations.

AI & CRISP-DM Integration

Low Code AI

Define the Project and Data Needs

- Select the fields or attributes to be extracted (e.g., well name, depth, production dates).
- Define the format for the final CSV output.

Train Low Code AI with Document Variations

- Document Collection
- Machine Learning in the application
- Training Iteration

Automated Data Extraction

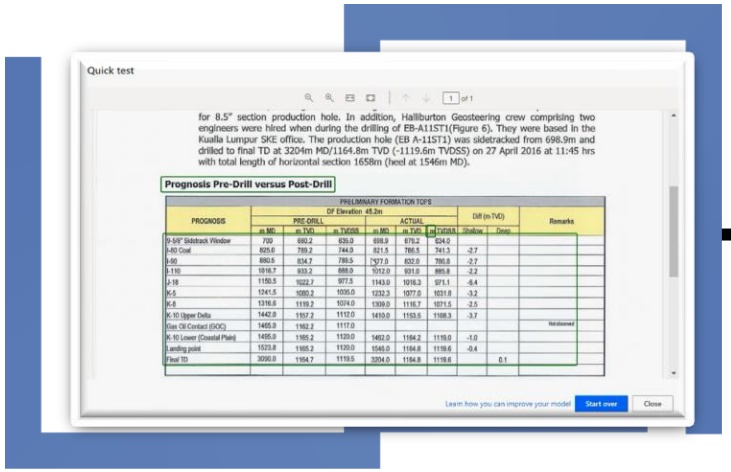
- The application uses the trained ML model to identify and extract key data fields.
- The application performs basic validation during extraction, checking that key fields are filled and formatted correctly

Data Storage

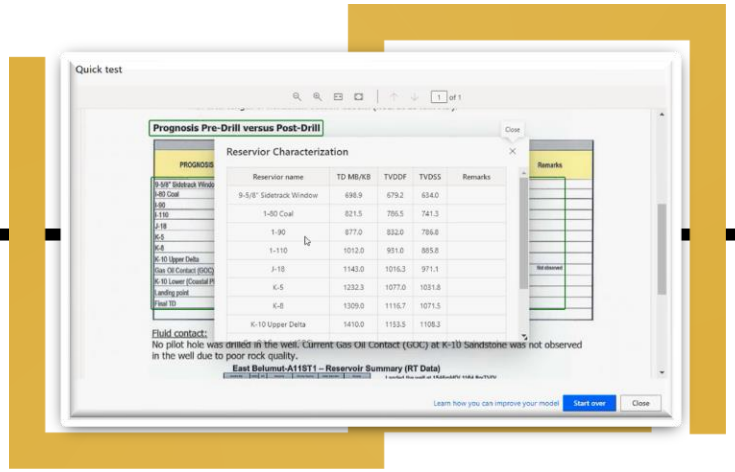
- The extracted data is automatically populated into a CSV file and/or database.

AI & CRISP-DM Integration

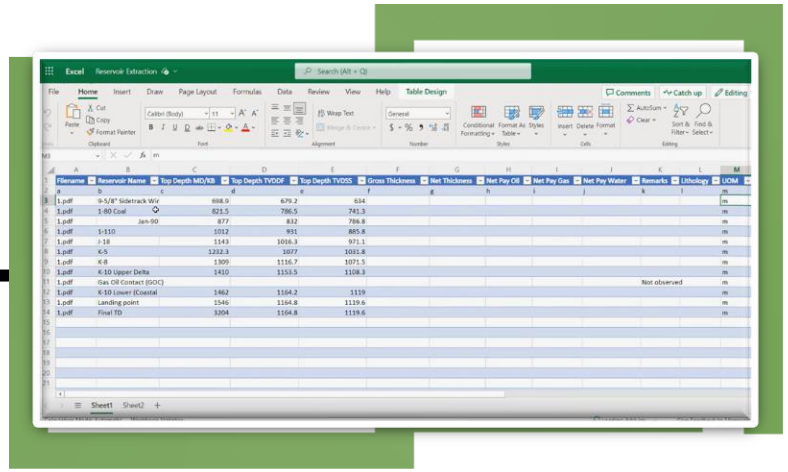
Low Code AI



Tagging the information in the unstructured report in **AI Builder**



Observed and QC the extracted information/data.



Export the information into CSV for further transformation and QC

Conclusion

The collaboration between PETRONAS and AEM Energy Solutions led to the successful automation of CCS Data Extraction and Transformation using the CRISP-DM framework.

- **Improved Data Accessibility:** Enabled 10x faster data retrieval, allowing project teams to access structured information more efficiently, supporting quicker decision-making.
- **Automation Efficiency:** CRISP-DM integration streamlined data extraction, reducing manual effort by utilizing Low Code AI, and increasing workflow efficiency.
- **Enhanced Data Quality & Recovery:** Approximately 261 new data entities and 4,540 attributes were accurately identified and integrated, ensuring reliable datasets for analysis.

The system is scalable and designed to support future CCS projects effectively.

"Empowering Data, Driving Efficiency—Securing the Future of CCS."