



ICTSS00120 - Artificial Intelligence Skill Set

Session 11: Understanding CRISP-DM

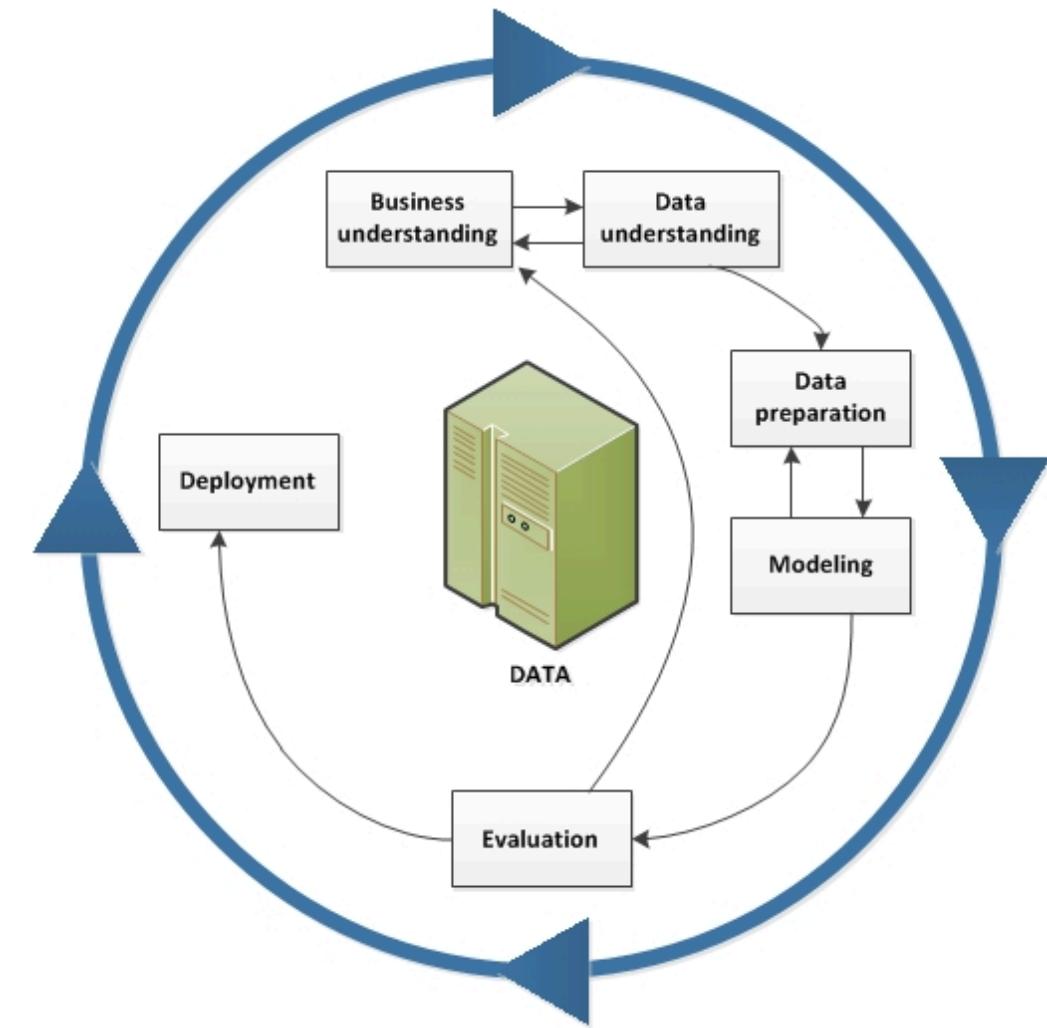
Lecturer: Jordan Hill

Learning Objectives

- Understand the CRISP-DM methodology and its importance in AI/ML projects.
- Apply Phase 1 – Business Understanding of CRISP-DM to a case study.
- Recognize the key activities in Phase 2 – Data Understanding.
- Explore how Phases 3, 4, & 5 integrate with prior learning.

What is CRISP-DM?

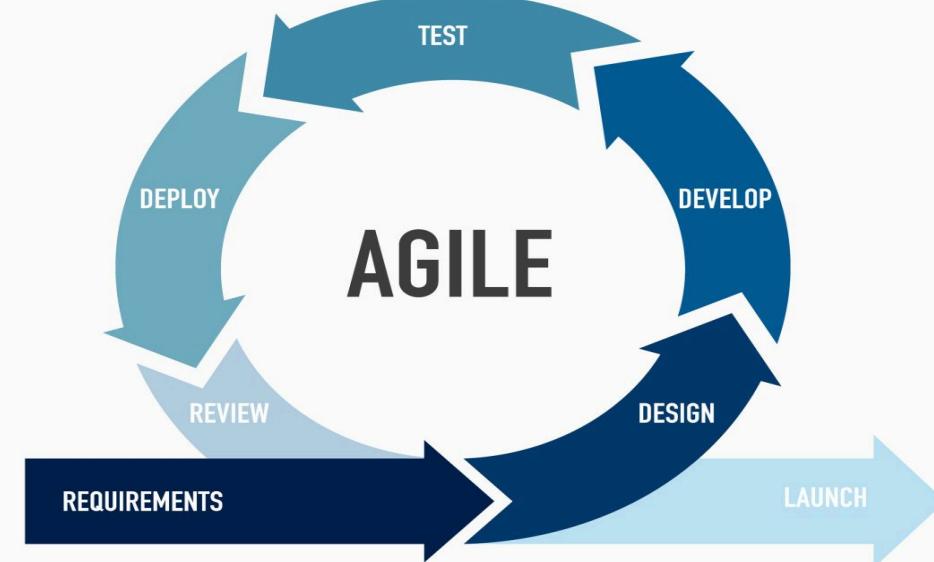
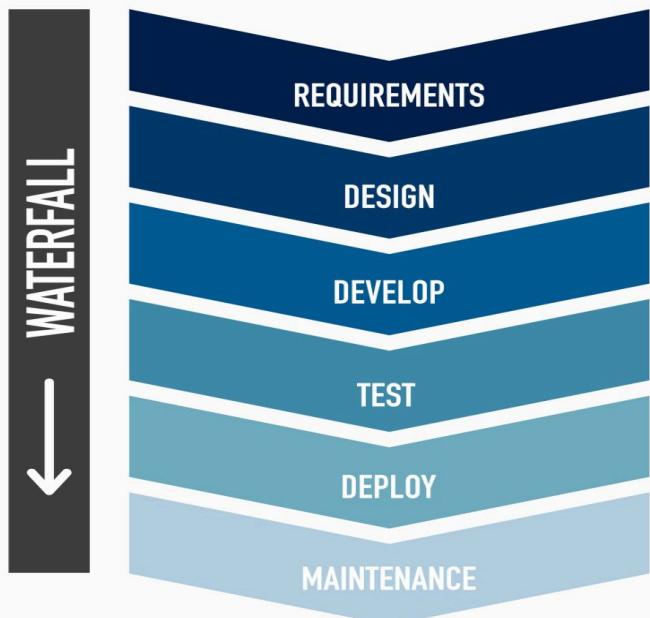
- CRISP-DM stands for **Cross-Industry Standard Process for Data Mining**.
- It provides a **structured approach** for planning and executing data mining projects.
- Widely used in industry for guiding data science and machine learning projects.



What do we already know about project management?

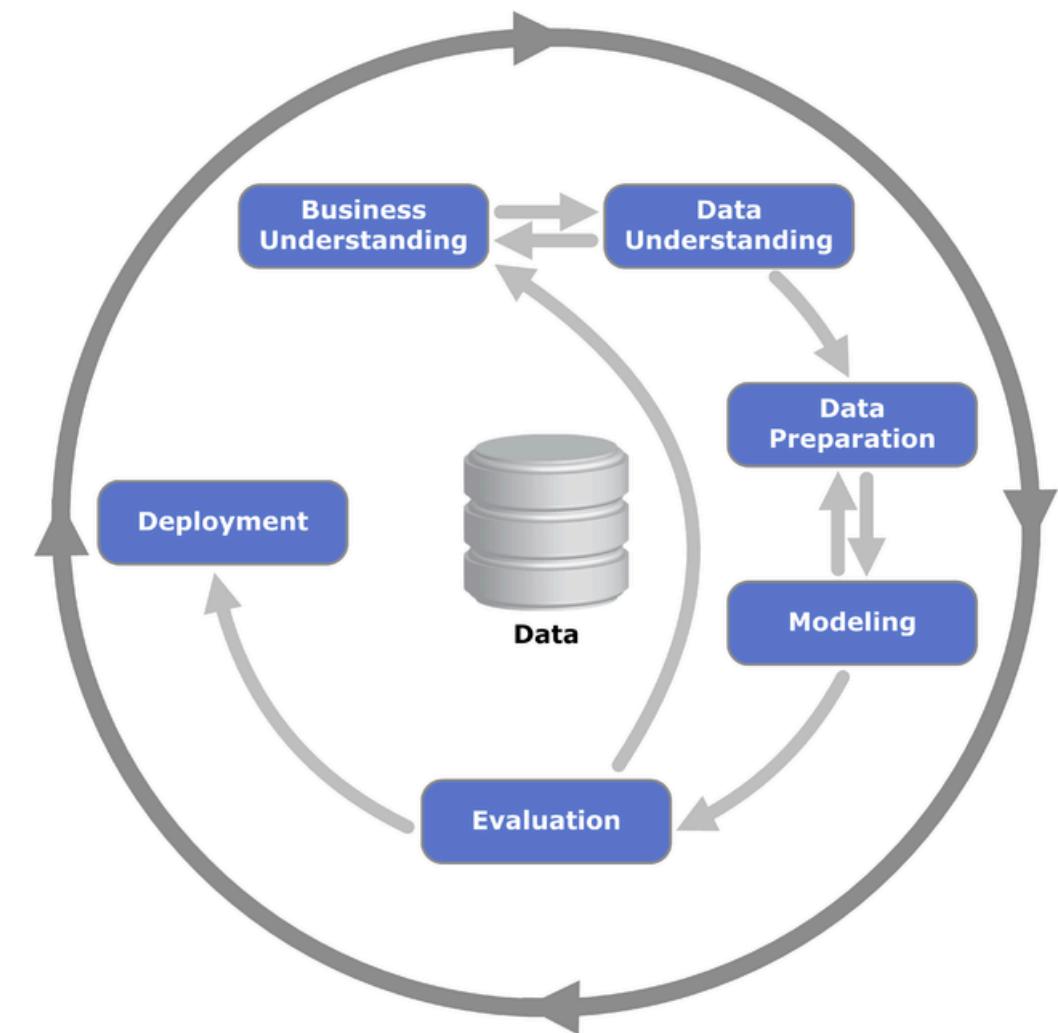
Agile? Waterfall?

AGILE vs WATERFALL



CRISP-DM Phases Overview

1. Business Understanding
2. Data Understanding
3. Data Preparation
4. Modeling
5. Evaluation
6. Deployment

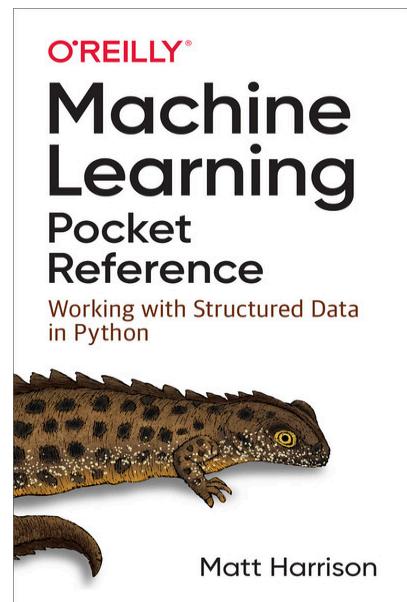


Why do we need this?

You will be expected to implement CRISP-DM throughout your final project

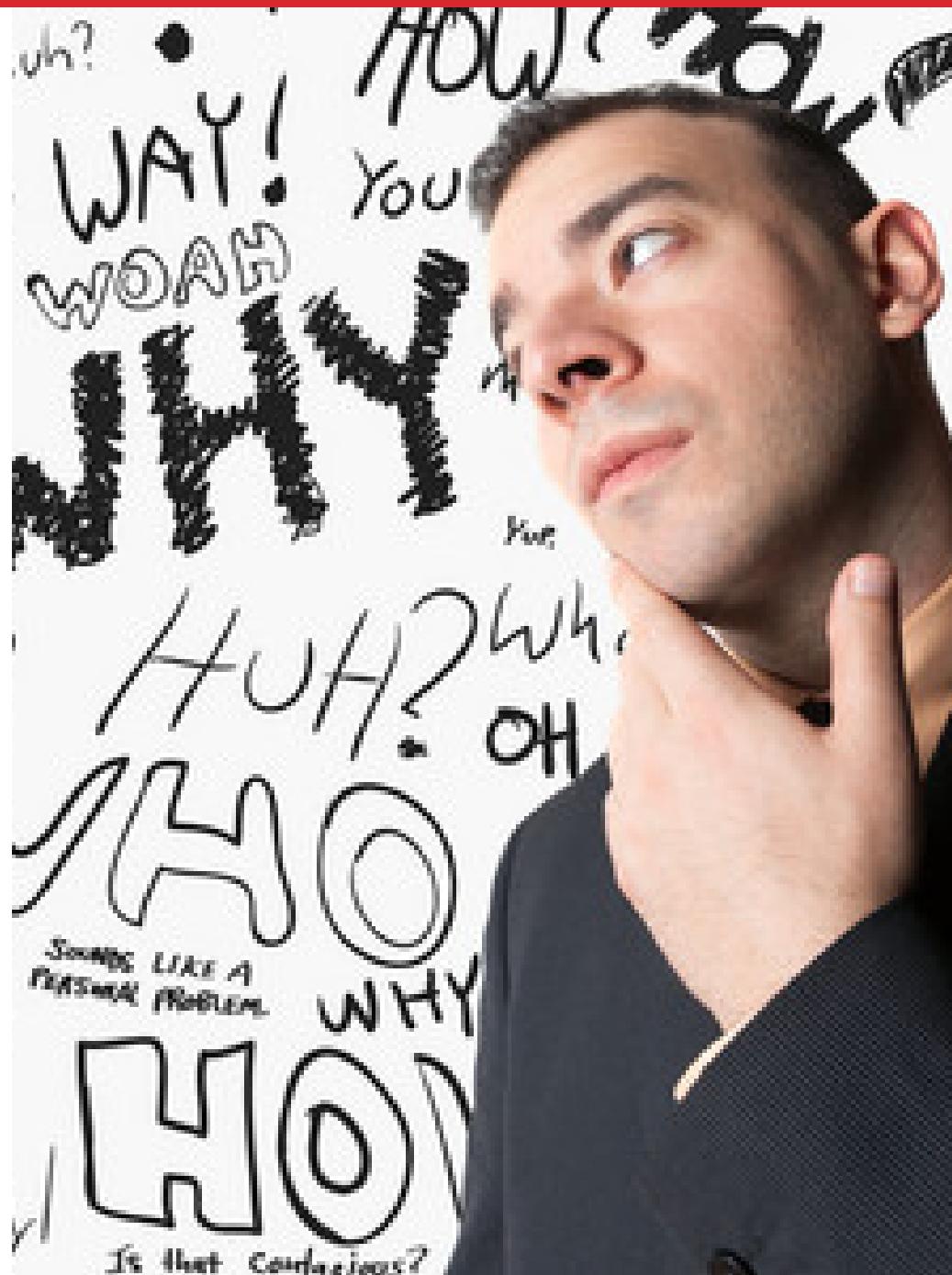
Today we will work through each stage in-depth

Reference Materials:



Phase 1: Business Understanding

- **Objective:** Understand the project objectives and requirements from a business perspective.
- **Key Steps:**
 - Determine Business Objectives
 - Assess the Situation
 - Establish Data Mining Goals
 - Produce Project Plan

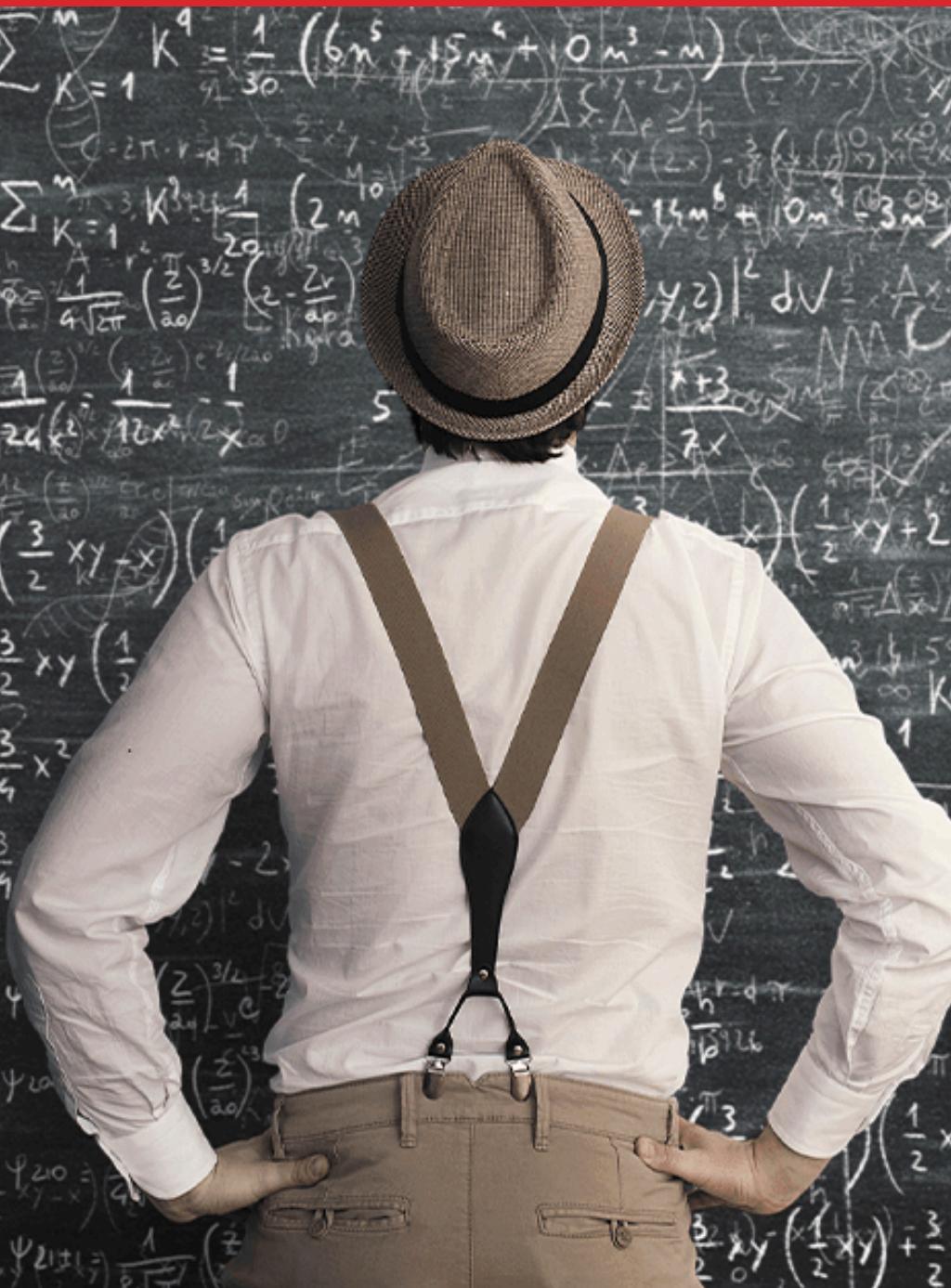


Activity: Define Business Objectives

- **In Groups:**
 - Choose a hypothetical business scenario or use your project idea.
 - **Identify** the main business objectives.
 - **Discuss** potential challenges and requirements.
 - **Share** your findings with the class.

Phase 2: Data Understanding

- **Objective:** Collect initial data and become familiar with it.
- **Key Steps:**
 - **Collect Initial Data**
 - **Describe Data**
 - **Explore Data**
 - **Verify Data Quality**





Importance of Data Collection

- **Accurate and relevant data** is critical for model success.
- **Data Quality Checks :**
 - Missing values
 - Outliers
 - Data consistency
- **Tools and Techniques :**
 - Data visualization
 - Statistical analysis

Phases 3, 4, & 5: Preparation, Modeling, Evaluation

Phase 3: Data Preparation	Phase 4: Modeling	Phase 5: Evaluation
<ul style="list-style-type: none">• Clean and format data for modeling.• Feature selection and engineering.	<ul style="list-style-type: none">• Select modeling techniques.• Build and test models.	<ul style="list-style-type: none">• Evaluate model performance.• Check if business objectives are met.



Integrating CRISP-DM with Prior Learning

- **Data Preprocessing Techniques** (from Week 2)
 - Applied during **Data Preparation**.
- **Algorithm Selection** (from Week 3)
 - Relevant in the **Modeling** phase.
- **Evaluation Metrics** (from Week 8)
 - Used in the **Evaluation** phase.

Worked Example: Applying CRISP-DM

- **Scenario:** Let's apply CRISP-DM to detect breast cancer!
- **Steps:**
 - Discuss business objectives.
 - Explore and prepare the dataset.
 - Choose and build a model.
 - Evaluate the model's performance.



[Kaggle Notebook](#)



Questions & Discussion

- Any questions about the CRISP-DM methodology?
- How can you apply CRISP-DM to your projects?
- What challenges might you face in each phase?

Next Week's agenda: Back to Transformers! Transformers and Final Project Preparation

For next week:

I *highly* recommend you try to watch this walkthrough by Andrej Karpathy (ex-OpenAI/ex-Tesla)

[**Andrej Karpathy: Let's build GPT: from scratch, in code, spelled out.**](#)

Check out the Notebook on Colab

[Github for video](#)

