CRISP-DM

How to break down a data modeling problem into an actionable solution

1

1

CRISP-DM

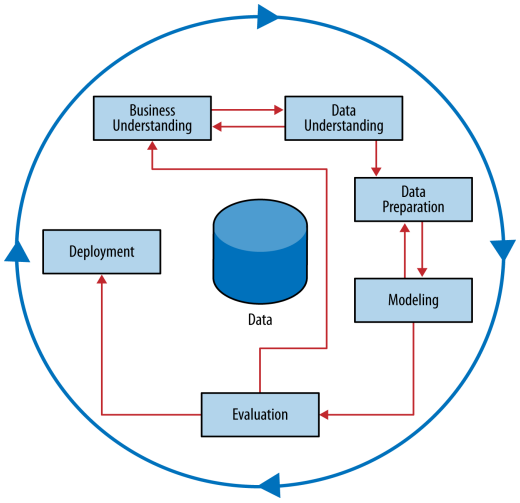
- An open standard process model that describes common approaches used by data mining experts. It is the most widely-used analytics model.
- Additional References
  - IBM SPSS Modeler CRISP-DM Guide
    - https://www.ibm.com/support/knowledgecenter/SS3RA7_sub/modeler_crispdm_ddita/modeler_crispdm_ddita-gentopic1.html
  - CRISP-DM 1.0
    - https://www.the-modeling-agency.com/crisp-dm.pdf
  - A Beginner's Guide to Industry Standard Process of Data Mining: CRISP-DM
    - https://medium.com/analytics-vidhya/a-beginners-guide-to-industry-standard-process-of-data-mining-crisp-dm-c1d7d50e57c3

2

2

CRISP-DM Phases

- Business Understanding
- Data Understanding
- Data Preparation
- Modeling
- Evaluation
- Deployment



3

3

CRISP-DM Tasks and Outputs

Business Understanding	Data Understanding	Data Preparation	Modeling	Evaluation	Deployment
Determine Business Objectives *Background* *Business Objectives* *Business Success Criteria* **Assess Situation** *Inventory of Resources* *Requirements, Assumptions, and Constraints* *Risks and Contingencies* *Terminology* *Costs and Benefits* **Determine DM Goals** *Data Mining Goals* *Data Mining Success Criteria* **Produce Project Plan** *Project Plan Initial Assessment of Tools and Techniques*	**Collect Initial Data** *Initial Data Collection Report* **Describe Data Data** *Description Report* **Explore Data Data** *Exploration Report* **Verify Data Quality** *Data Quality Report*	**Select Data** *Rationale for Inclusion/ Exclusion* **Clean Data** *Data Cleaning Report* **Construct Data** *Derived Attributes* *Generated Records* **Integrate Data** *Merged Data* **Format Data** *Reformatted Data* *Dataset* *Dataset Description*	**Select Modeling Techniques** *Modeling Technique* *Modeling Assumptions* **Generate Test Design** *Test Design* **Build Model Parameter** *Settings Models* *Model Descriptions* **Assess Model** *Model Assessment* *Revised Parameter Settings*	**Evaluate Results** *Assessment of Data Mining Results Business Success Criteria Approved Models* **Review Process** *Review of Process* **Determine Next Steps** *List of Possible Actions* *Decision*	**Plan Deployment** *Deployment Plan* **Plan Monitoring and Maintenance** *Monitoring and Maintenance Plan* **Produce Final Report** *Final Report* *Final Presentation* **Review Project** *Experience Documentation*

Generic tasks (blue and bold) and *outputs (italic)*

https://www.the-modeling-agency.com/crisp-dm.pdf

4

4

Business Understanding

Description

- Thoroughly understand, from a business perspective, what the customer really wants to accomplish.
  - Business objectives
  - Assess Situation
  - Determine DM Goals
  - Produce Project Plan

Appleton Example

- Turn prospects into customers.
- Prospects need more information to decide.
- Accurately and quickly determine interest rate.

5

5

Data Understanding

Description

- Collect and assess the data that has been collected and/or the data that can be collected.
  - Collect Initial Data
  - Describe Data Data
  - Explore Data Data
  - Verify Data Quality

Appleton Example

- What is currently being collected?
- Is any more data readily available?
- What other data could be collected or aggregated?

6

6

Data Preparation

Description

- Decide on the data to be used and clean and format into a viable data infrastructure
  - Select Data
    - Rationale for Inclusion/ Exclusion
  - Clean Data
  - Construct Data
  - Integrate Data
  - Format Data

Appleton Example

- What data should be used and prep for modeling.
  - Drop features
  - Missing values
  - Labels to numeric
  - Transform data
  - Merge, concat, etc.

7

7

Modeling

Description

- Select the modeling technique
- Do you know the historical target?
  - Unsupervised
  - Supervised
- Are you predicting a numerical value?
  - Regression
    - Accuracy, parsimonious, or significant features
- Are you predicting a decision?
  - Classification

Appleton Example

- Trying to predict **Interest Rate**
  - Create most accurate model (R-Squared)
  - Parsimonious and significant features (backwards elimination)

8

8

Evaluation

Description

- Assess the degree to which the model meets the business objectives.
- Determine if there is some business reason why this model is deficient.
  - Are the missing values for important variables?
- Are the values that are missing likely to be larger or smaller than the values we have?

Appleton Example

- Is a model with R^2=0.521 accurate enough to make decisions?
  - If not, then evaluate what additional features can be collected.
- Do the significant features meet the business objective? ***Is it explainable?***

9

9

Deployment

Description

- Putting the model into operation.
- Produce plan for deployment.
- Is it feasible to put into production?

Appleton Example

- How can the model be deployed so that a customer gets an immediate answer?
  - Web-based solution

10

10

Summary

CRISP-DM

- Focuses data modeling on solving a business problem.
- Forces the process to identify data that is pertinent to the business problem.
- Enables the creation of clean and transformed data prior to modeling.
- Established procedures and best practices for modeling, evaluation, and deployment.

11

11

Credits

- Presentation template by SlidesCarnival
- Illustrations by Undraw.co

12

12