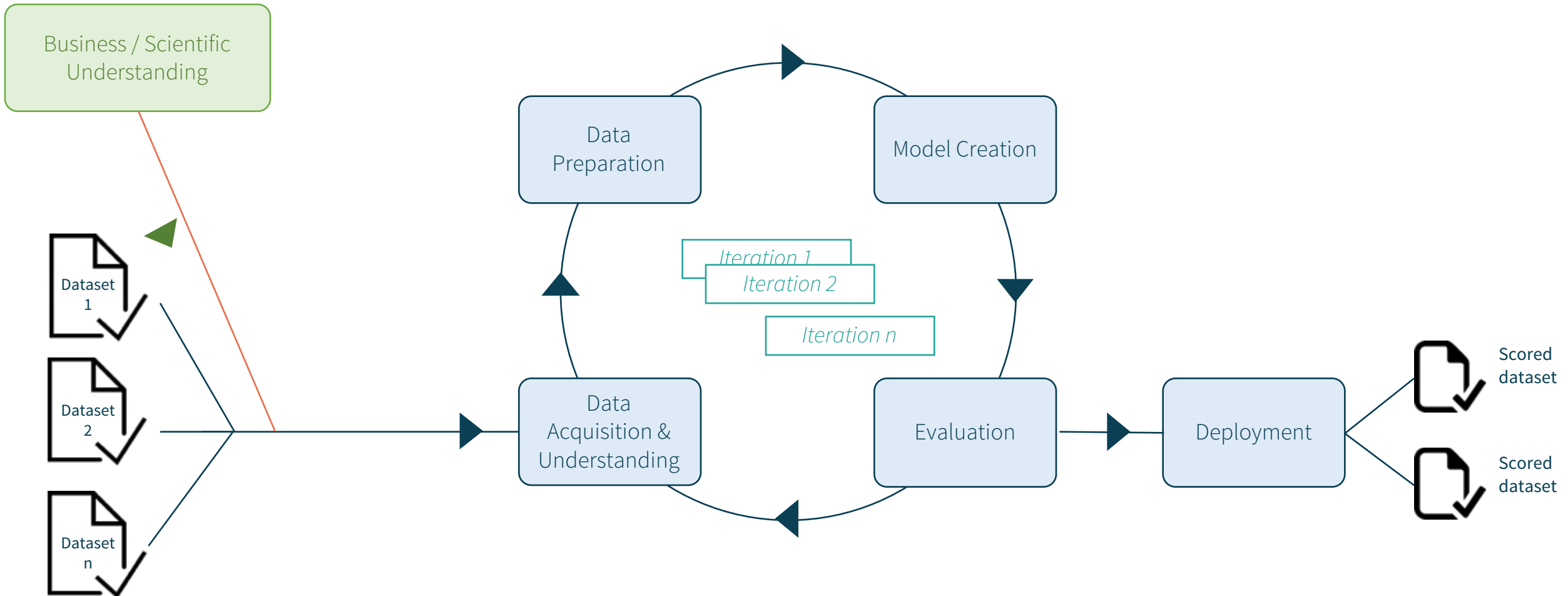




DSS + Stan  
Bayesian Data Science

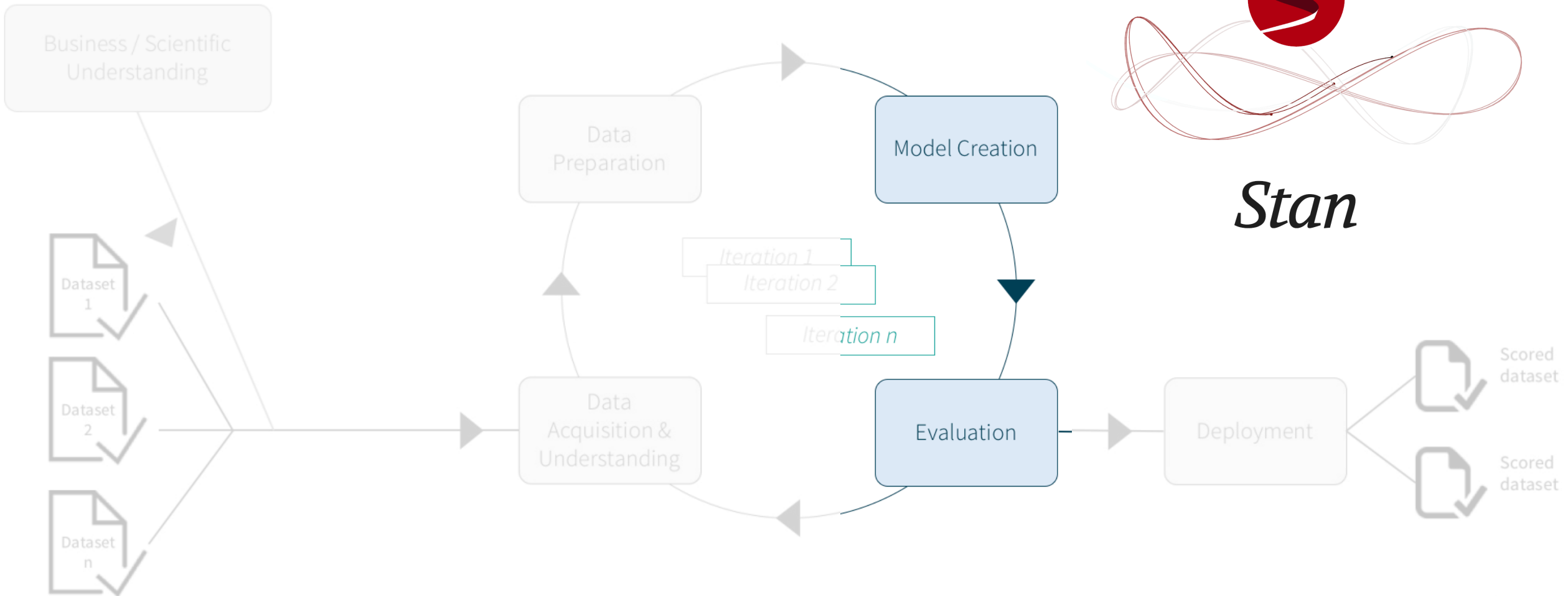
# A data science workflow



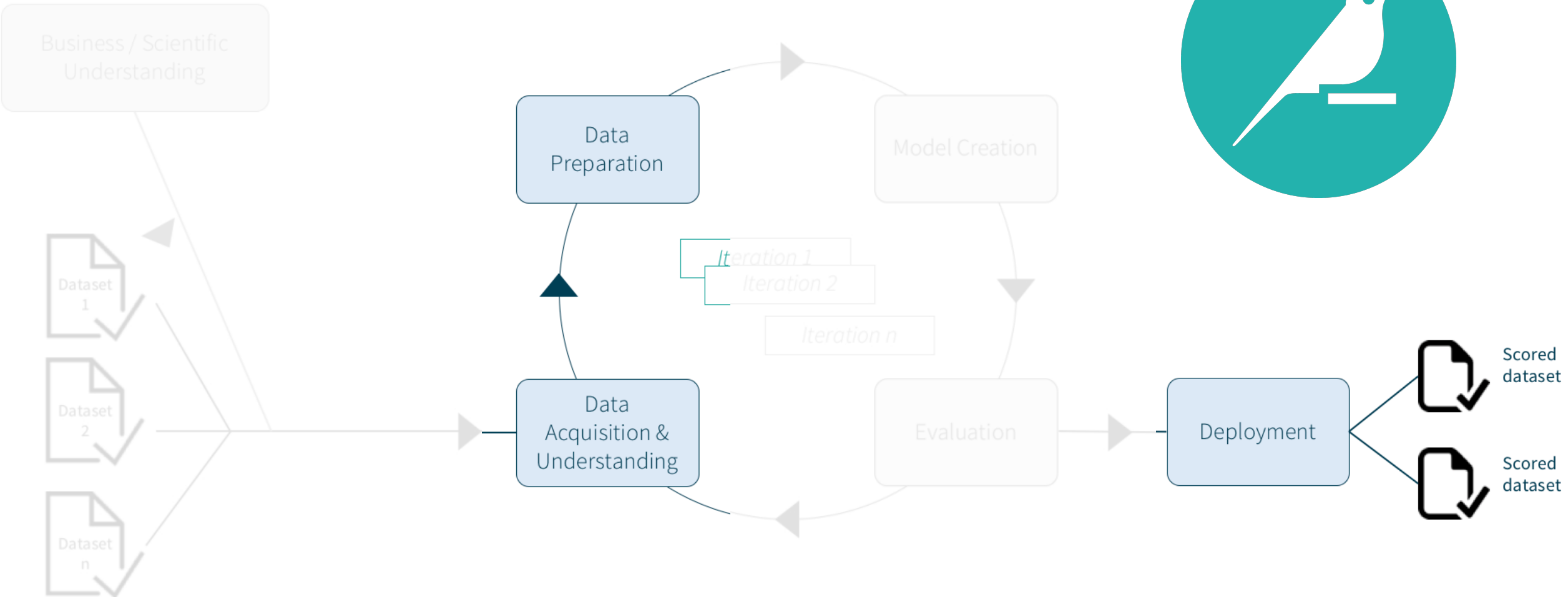
Adapted from the CRISP-DM methodology



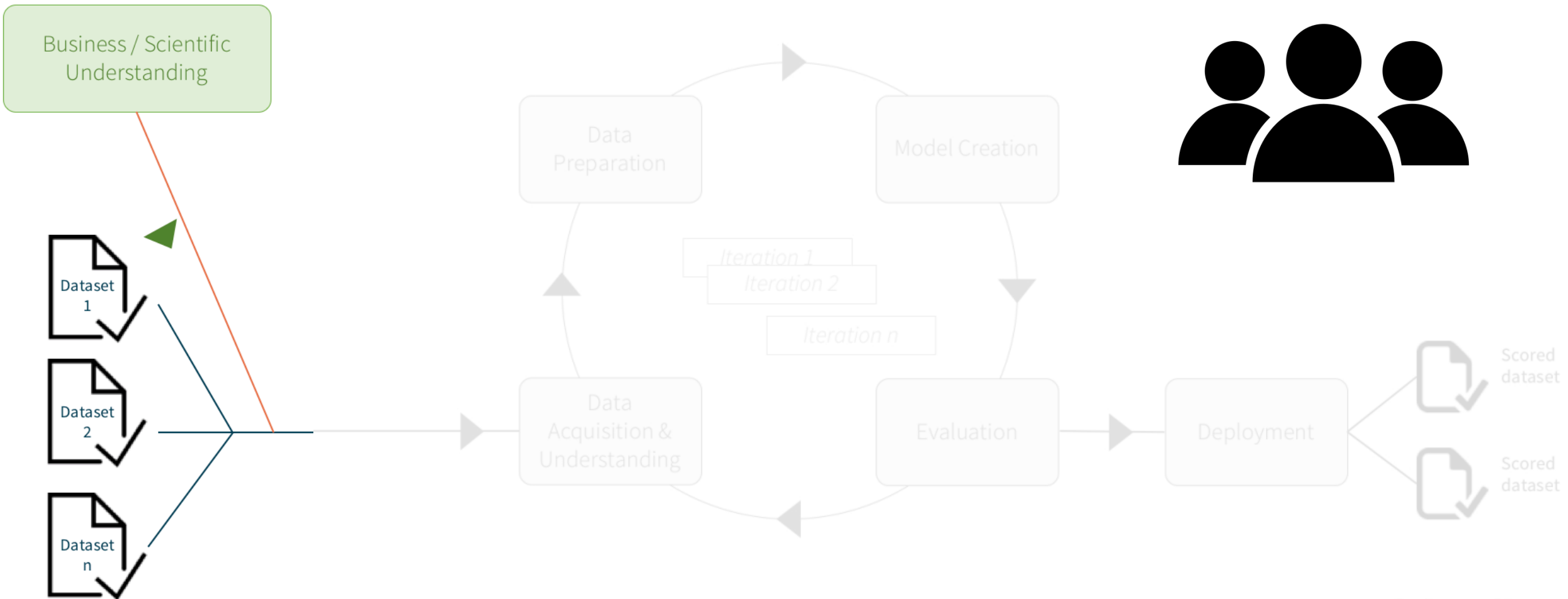
# A data science workflow



# A data science workflow

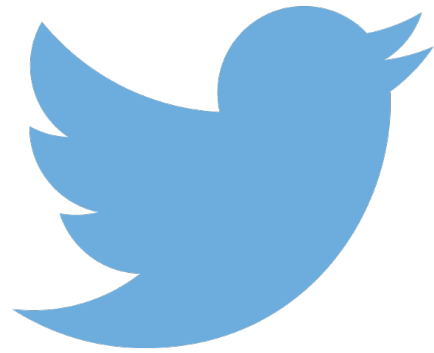


# A data science workflow



# Today's Demo

Topic Modeling of Tweets



+



+



# Topic Modeling

# Topic Modeling

A brief intro

Document



Topics

News

Pop culture

Marketing

Words

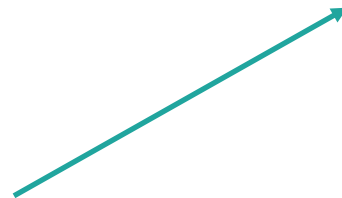
Product

Report

Award

Trump

Video

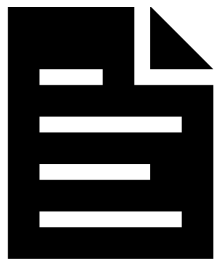




# Topic Modeling

A brief intro

Document



Topics

News

Pop culture

Marketing

Words

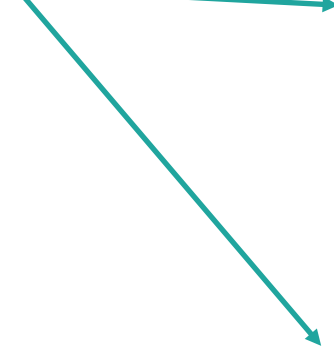
Product

Report

Award

Trump

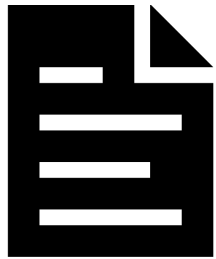
Video



# Topic Modeling

A brief intro

Document



Topics

News

Pop culture

Marketing

Words

Product

Report

Award

Trump

Video



# Bayesian Topic Modeling

Document

Topics

Words



20%

News

Product

70%

Pop culture

Report

10%

Marketing

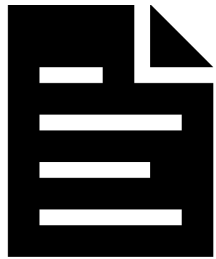
Award

Trump

Video

# Bayesian Topic Modeling

Document



Topics

News

Pop culture

Marketing

Words

5% Product

5% Report

40% Award

20% Trump

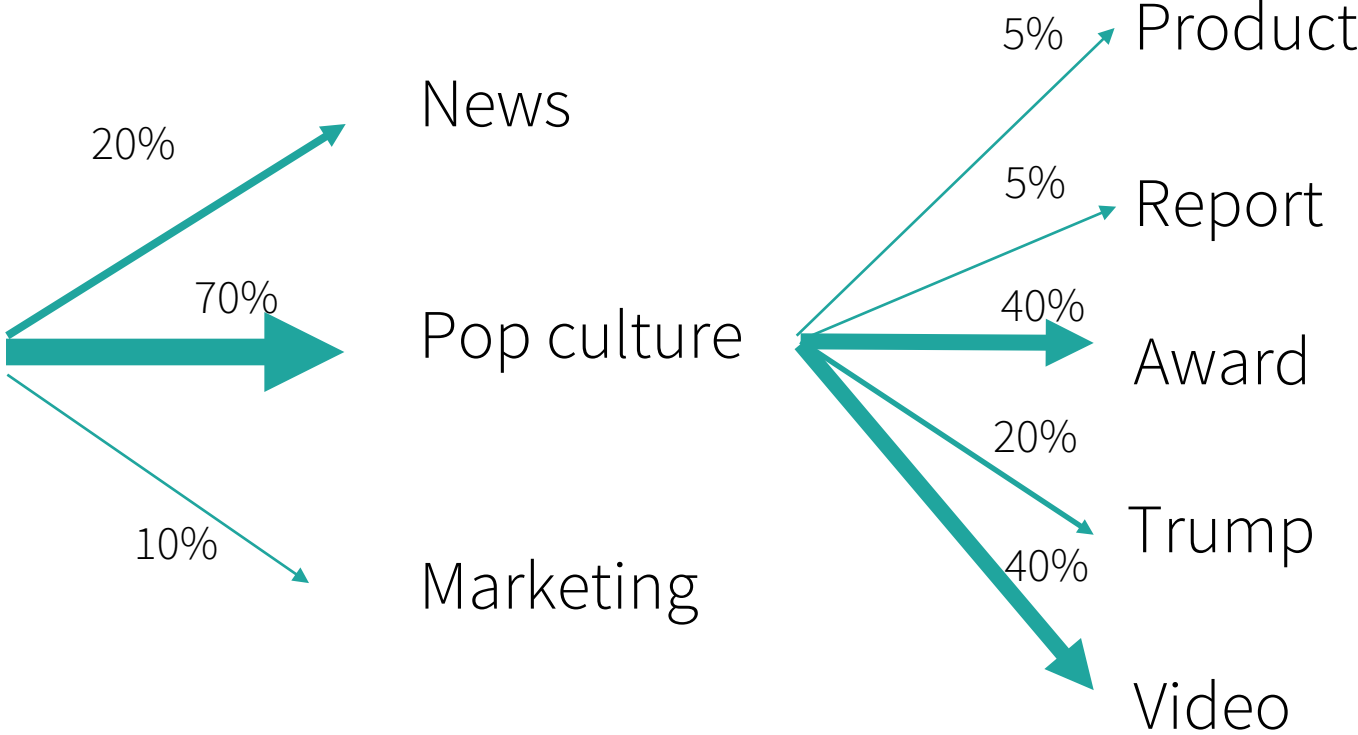
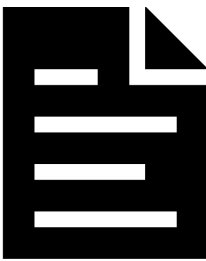
40% Video

# Bayesian Topic Modeling

Document

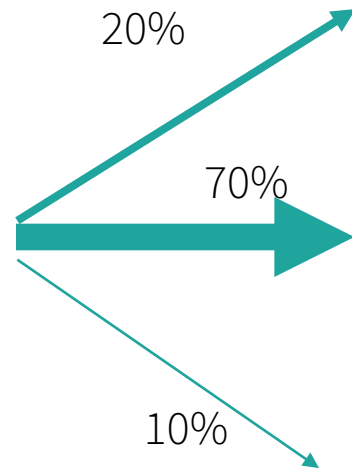
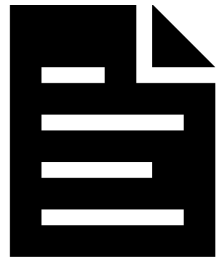
Topics

Words



# Topics are “hidden”

Document



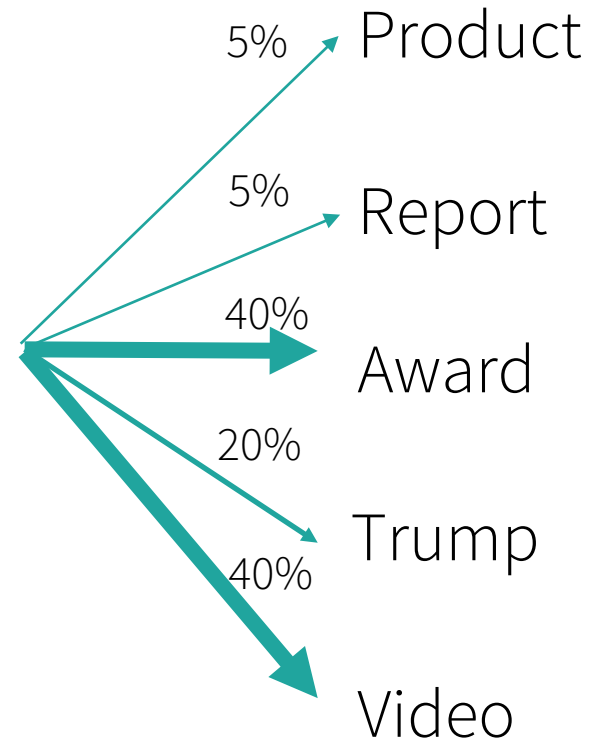
Topics

Topic 1

Topic 2

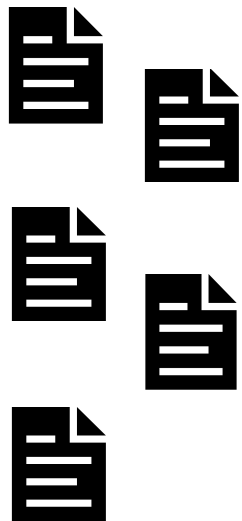
Topic 3

Words



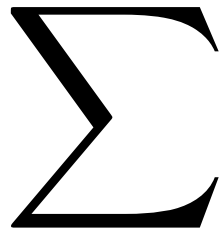
# Workflow

Documents



*Tweets*

Algorithm



*Latent  
Dirichlet  
Allocation*

Outputs

*Distribution of topics  
for each document*

*Distribution of words  
for each topic*

To DSS!