# UNIVERSITY of WASHINGTON

## Information School
### UNIVERSITY of WASHINGTON

# The modern product manager's toolkit!
## Sentiment Analysis and Topic Modeling

Rahul Zende, Anjali Singh, Hrishika Shetty, Bhagyashri Nivdunge

**smartsheet**

## Problem Statement

➢ How to decompose customer feedback into a set of topics, with no prior knowledge?

➢ How do we understand the customers' sentiments behind any and all feedback, with no prior labels?

## Team Goal

Our team goal is to empower our employers take data-driven business decisions

➢ Determine the most used Smartsheet features
➢ Determine the sentiments for them
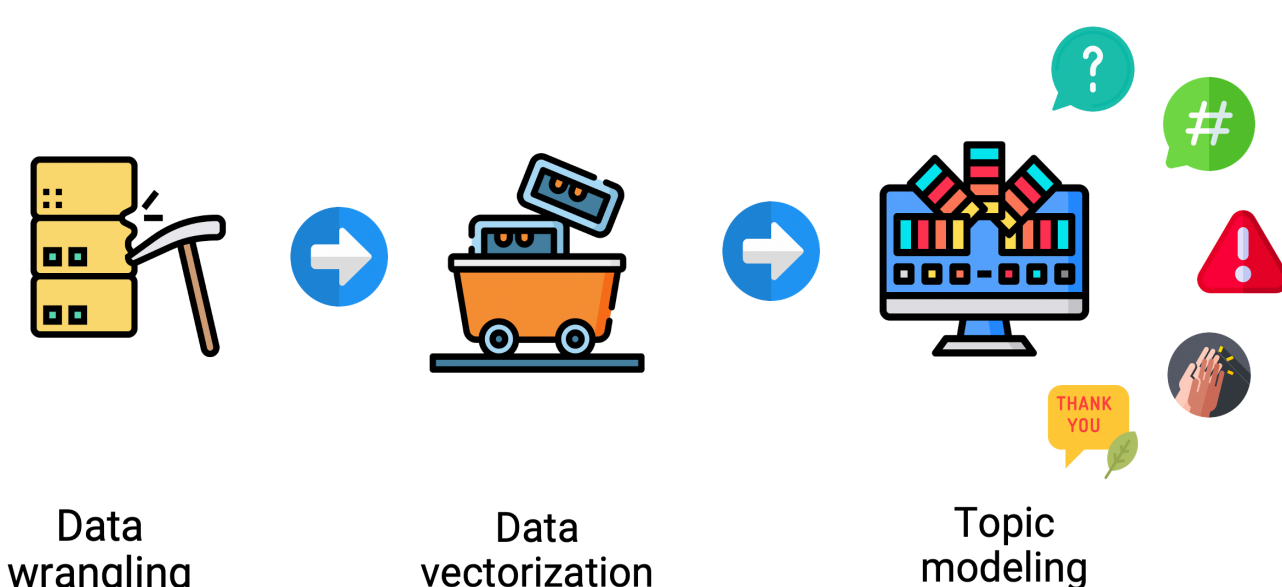➢ Visualize and recommend solution based on the analysis

## Proposed Solution
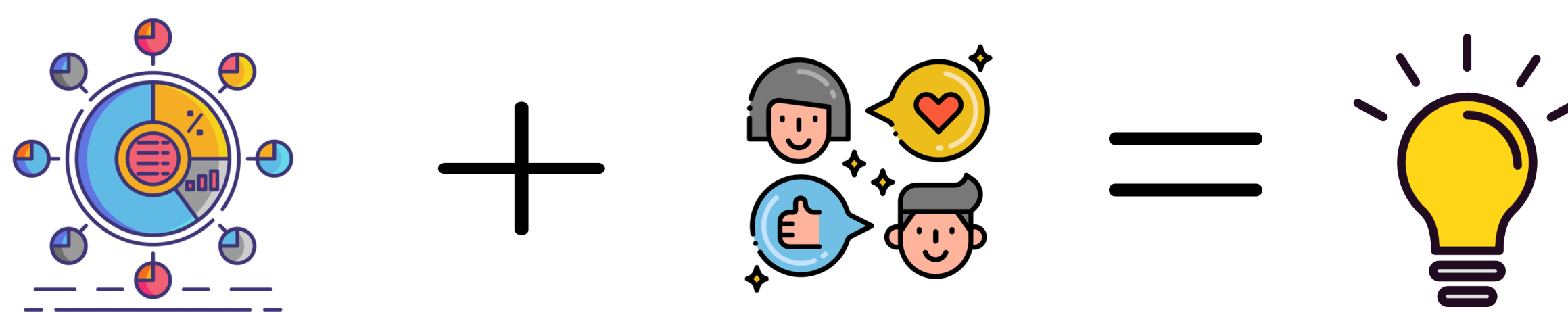


Topic Modeling  +  Sentiment Analysis  =

## Topic Modeling



Data wrangling   Data vectorization   Topic modeling

➢ Use cross-validation to finalize the optimum number of topics derived
➢ List the 10 most frequently occurring words in each topic
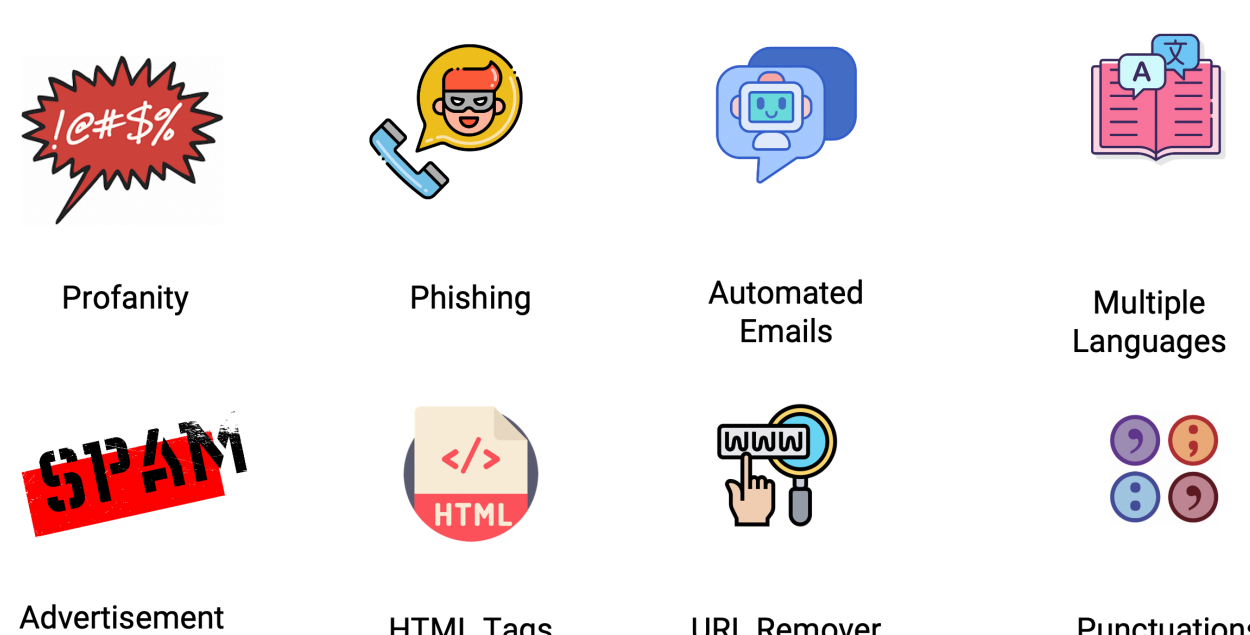➢ Select a model that offers best balance of interpretability & functionality



### Data Sources
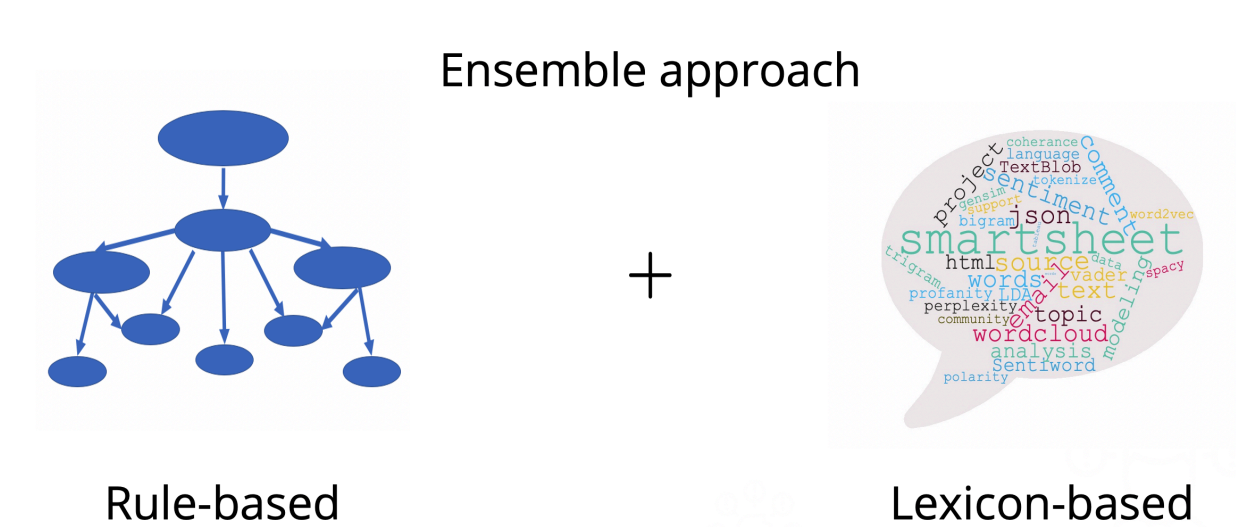
• **Community Platform Data**
• **Customer Support Data**

> 107k + 64k customer feedback comments each

> Different data formats (CSV and three JSON files)

### Challenges in cleaning data



Profanity    Phishing    Automated Emails    Multiple Languages

Advertisement    HTML Tags    URL Remover    Punctuations

## Sentiment Analysis

Ensemble approach



Rule-based  +  Lexicon-based

➢ Unsupervised learning (no labels)
➢ Verbiage of feedback is difficult to classify
➢ Optimizing tools and libraries to achieve best results
➢ Solution : Rule-based + Lexicon-based approach



## Final Approach

➢ 3 topic models for deriving underlying themes in data
➢ 3 sentiment analysis libraries to perform a combination of lexicon-based & rule-based classification
➢ Combine these in Tableau dashboards to get a holistic overview of customer feedback
➢ Focused on identifying negative customer feedback & provided insights and key takeaways for future product roadmap