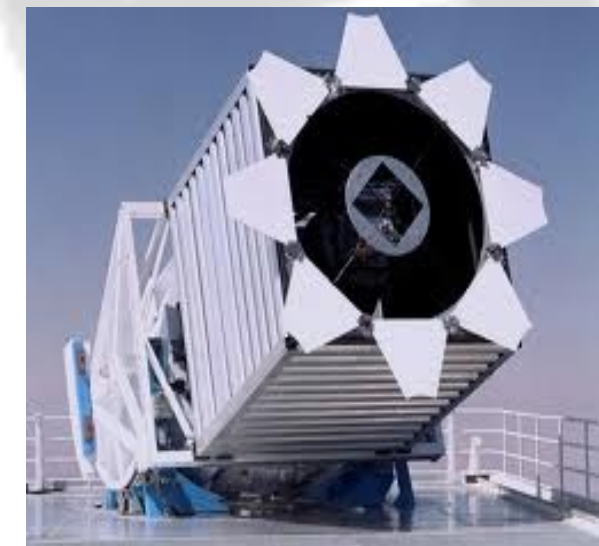


Interactive Data Exploration

- Human-in-the-loop applications that search big datasets to discover interesting information
- A long-running, multi-step process with end-goals not stated explicitly



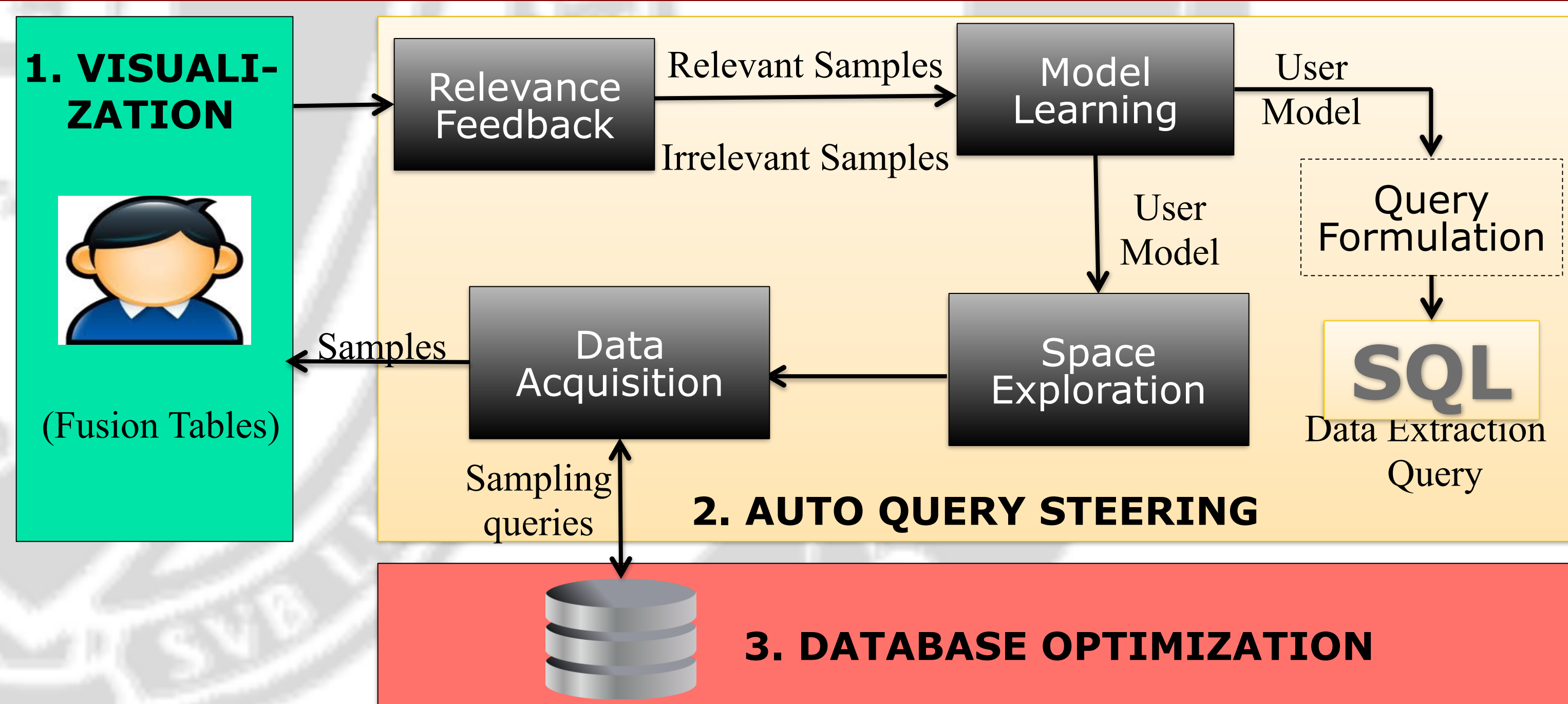
Medical Applications



Scientific Applications



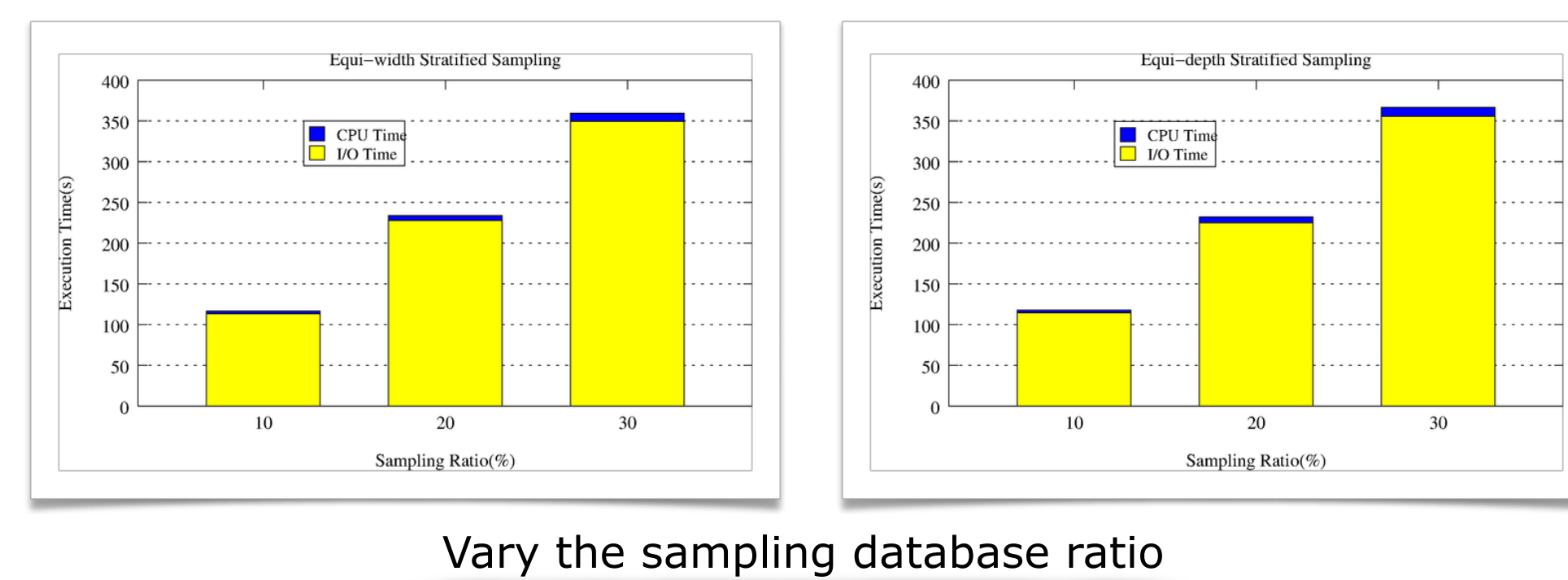
Framework Overview and Challenges



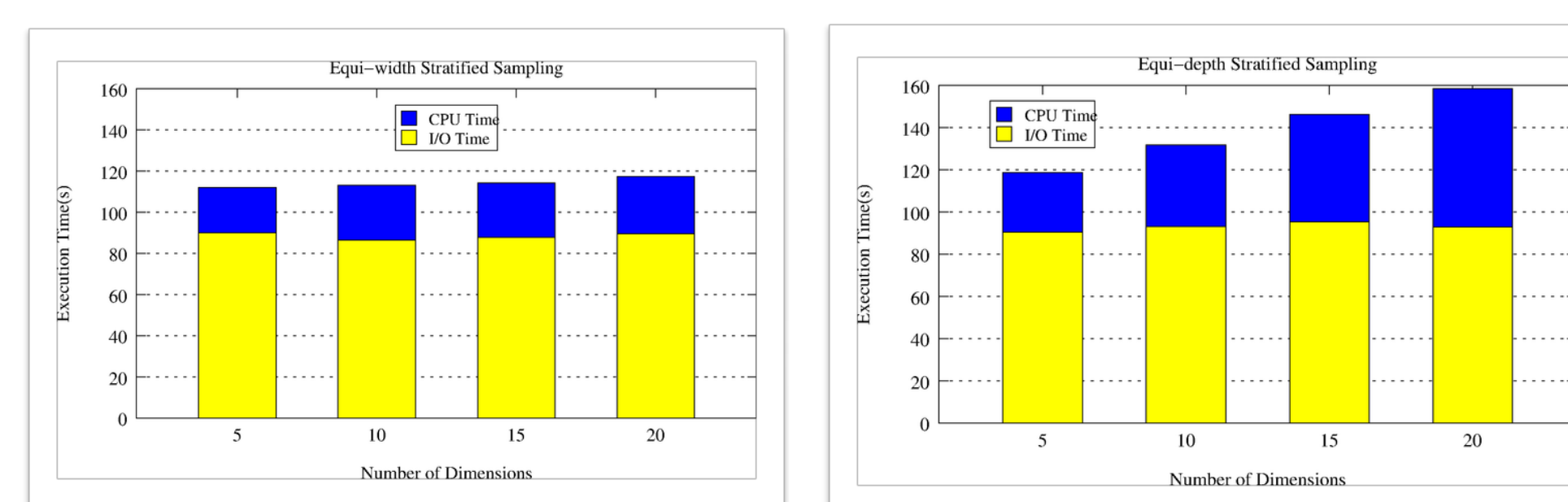
- Effective Data Exploration:** Which data samples to show to the user?
 - Unknown user interests
- Efficient Sample Acquisition:** How to minimize acquisition cost?
 - High sample acquisition cost on big data sets
 - Accuracy vs efficiency trade-off
 - Model learning & sample acquisition need to be coupled

Initial Sampling

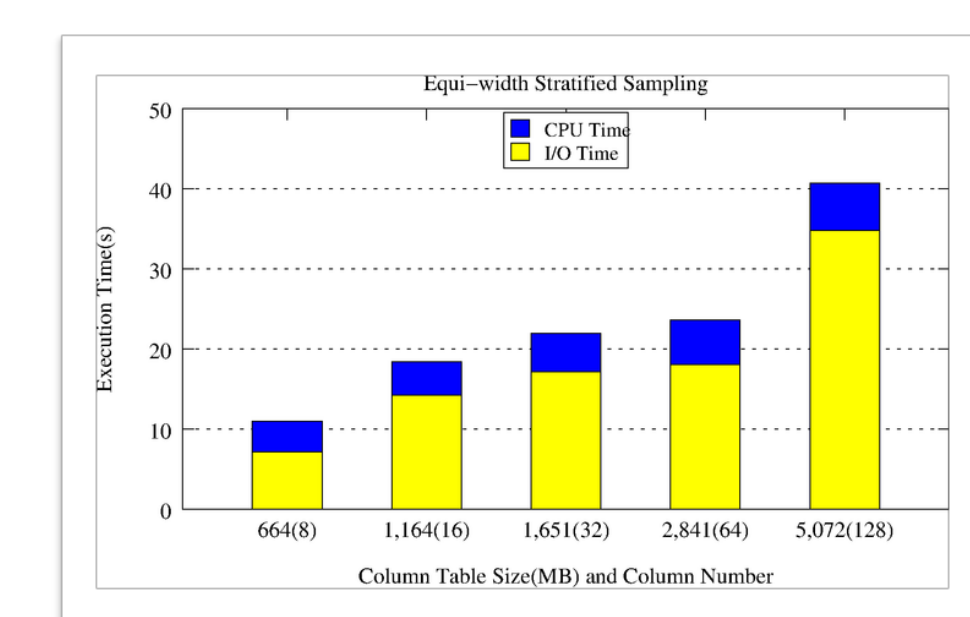
- Initial Sampling**
 - Select samples in the first round
 - Before any user feedback information
- Compare Equi-width and Equi-depth Stratified Sampling**
 - Change the size of the sampling database
 - Use column tables with different number of columns
 - Perform sampling within different dimensions



Vary the sampling database ratio

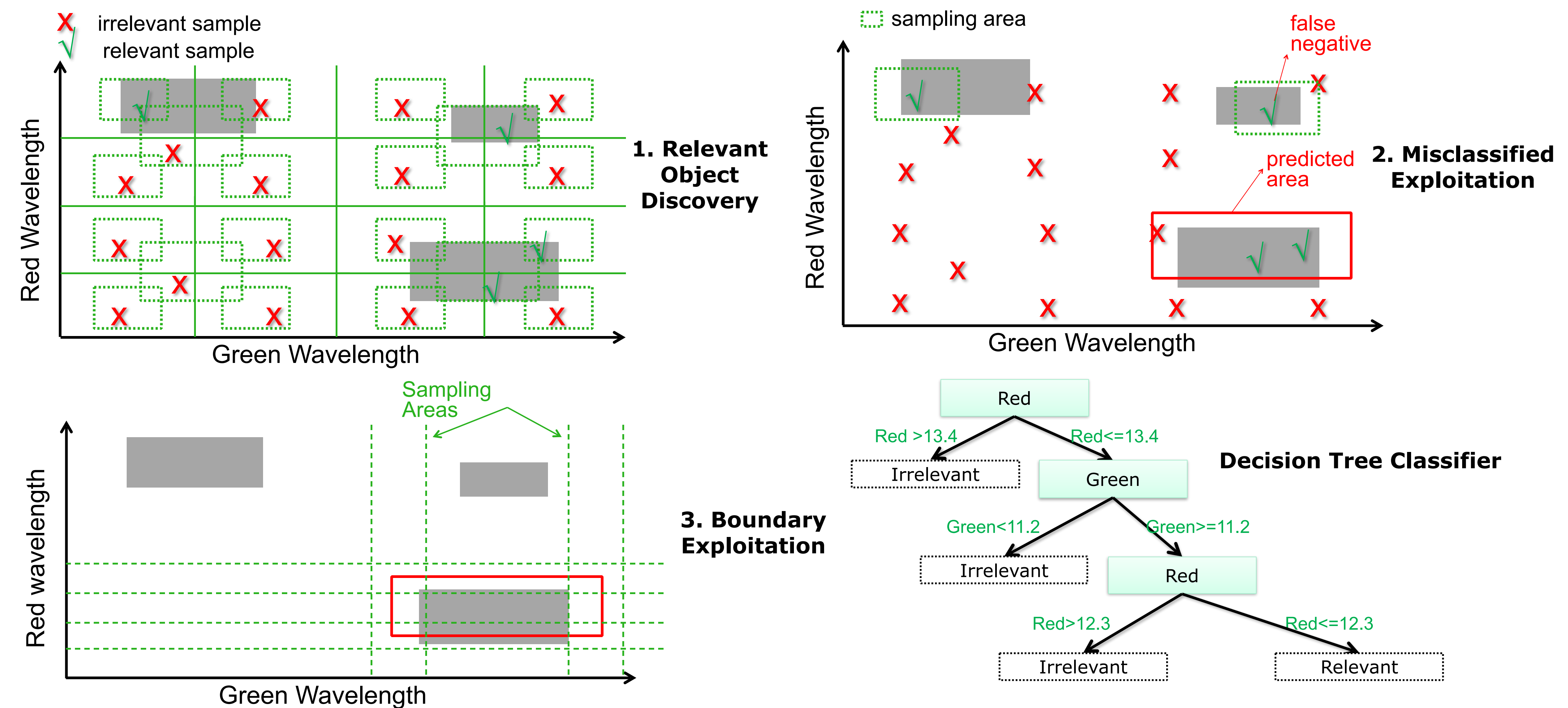


Vary the number of dimensions



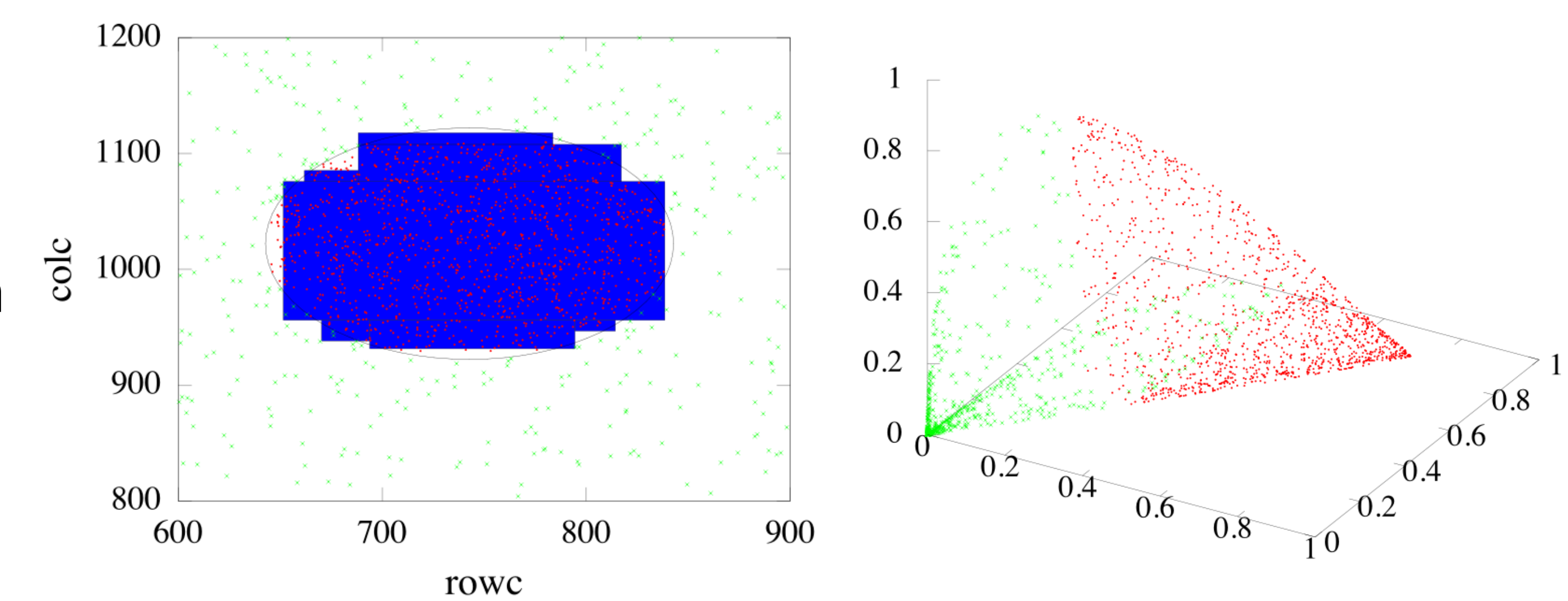
Vary the column table

Decision-tree-based Linear Pattern Discovery



SVM-based Non-linear Pattern Discovery

- Support Vector Machine
 - Large margin classifier
 - Kernel tricks
 - A strong math foundation
- Space Exploration
 - Active learning
 - Utilizing unlabeled data



Front-end Demonstration

