

Interactive Construction of Exploratory Queries for Highly Imbalanced Data



Xiaoyu Ge, Xiaozhong Zhang, Panos K. Chrysanthis

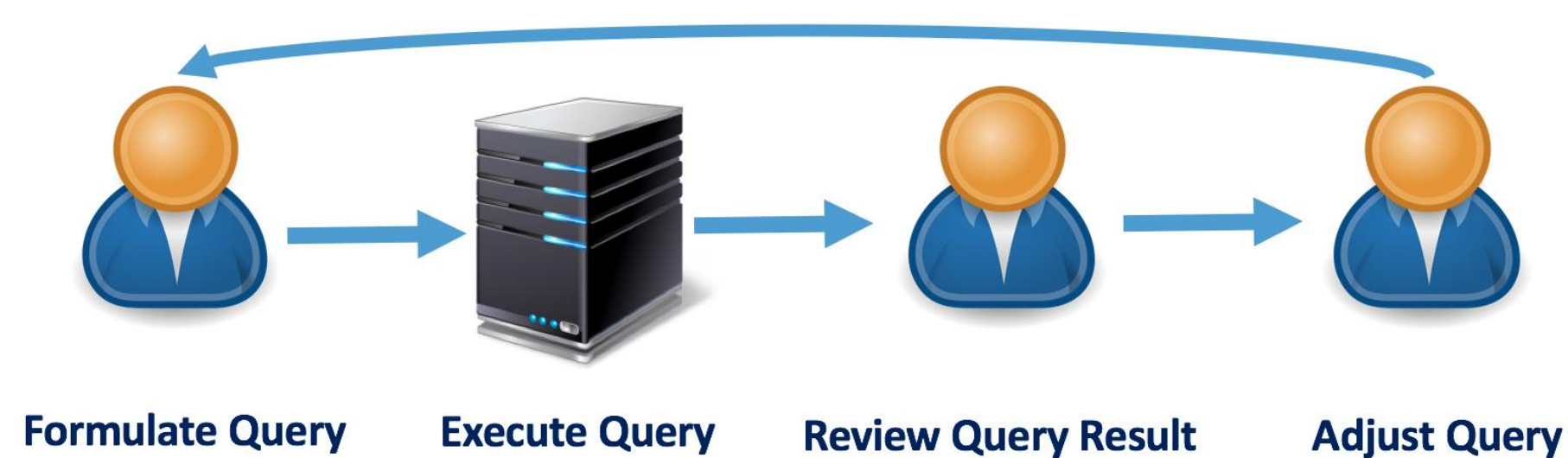


Motivation

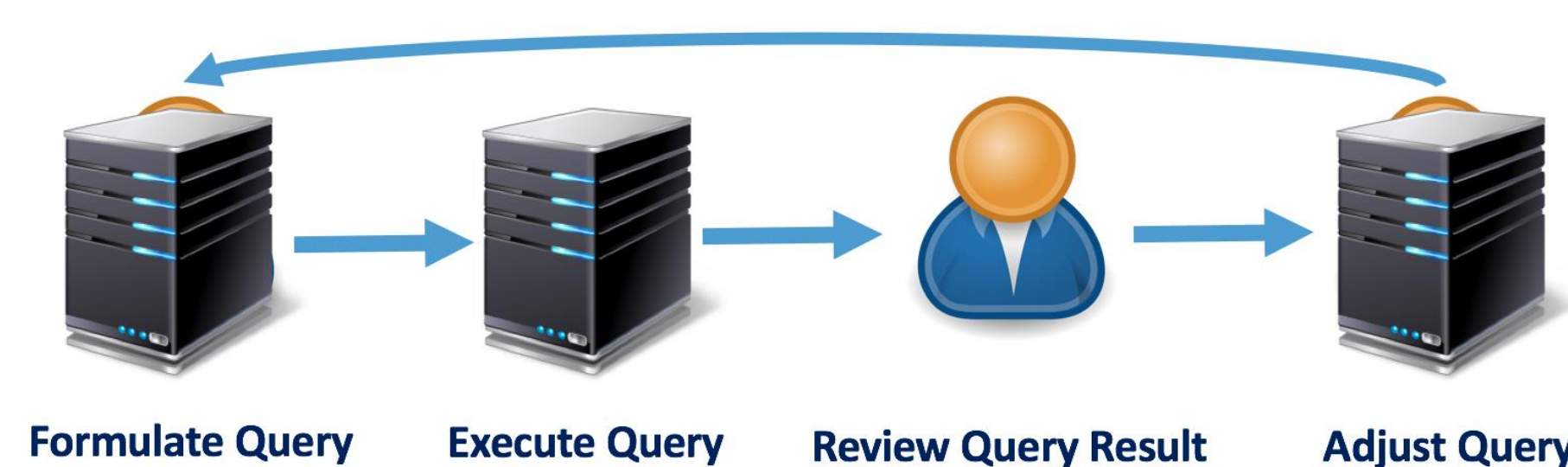


- ❖ Need to explore large volume of data for valuable insights.
- ❖ Users are unfamiliar of SQL and underlying database.
- ❖ Examples including scientific computing, financial analysis and etc...

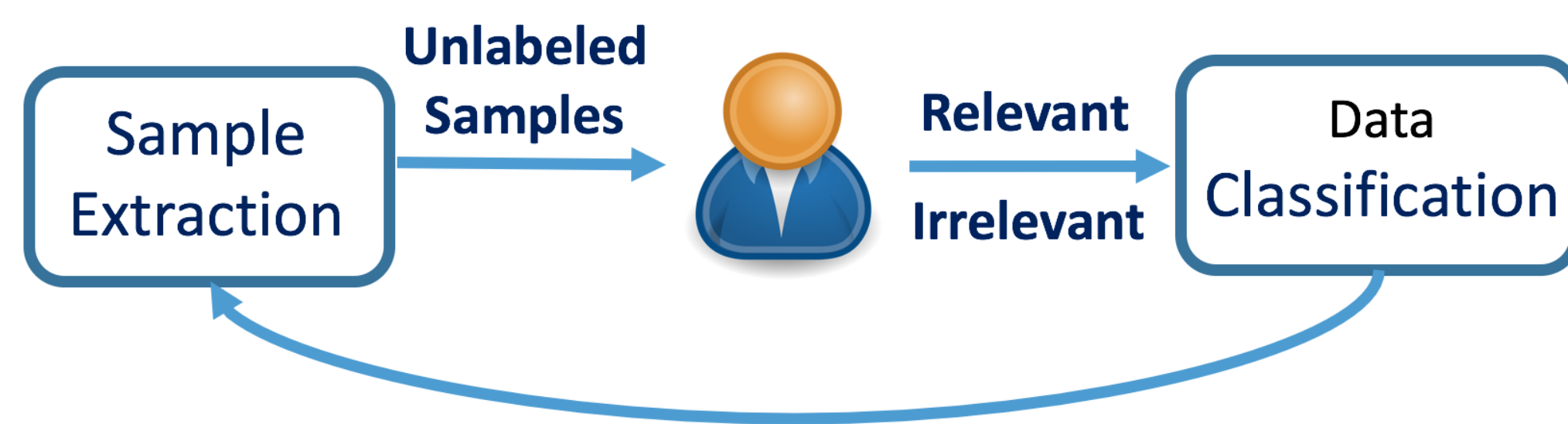
Query-from-Examples (QFE)



❖ Manual Data Exploration is long, imprecise and labor intensive!

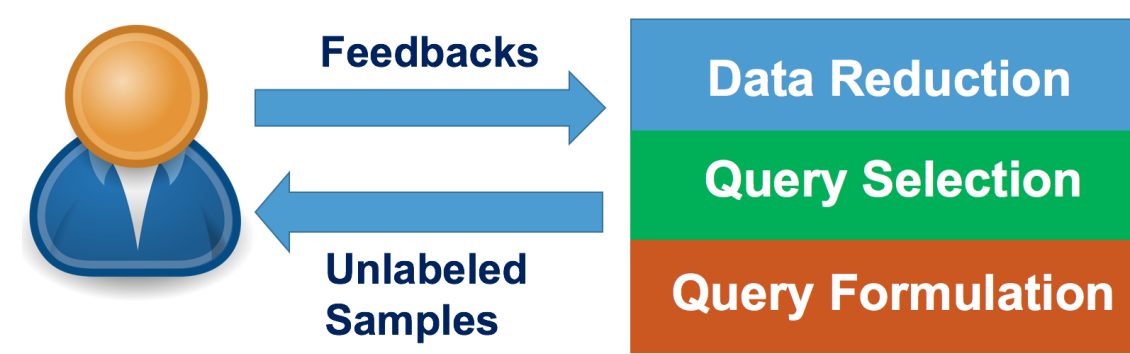


❖ Automatic Data Exploration is more precise and needs less labor!



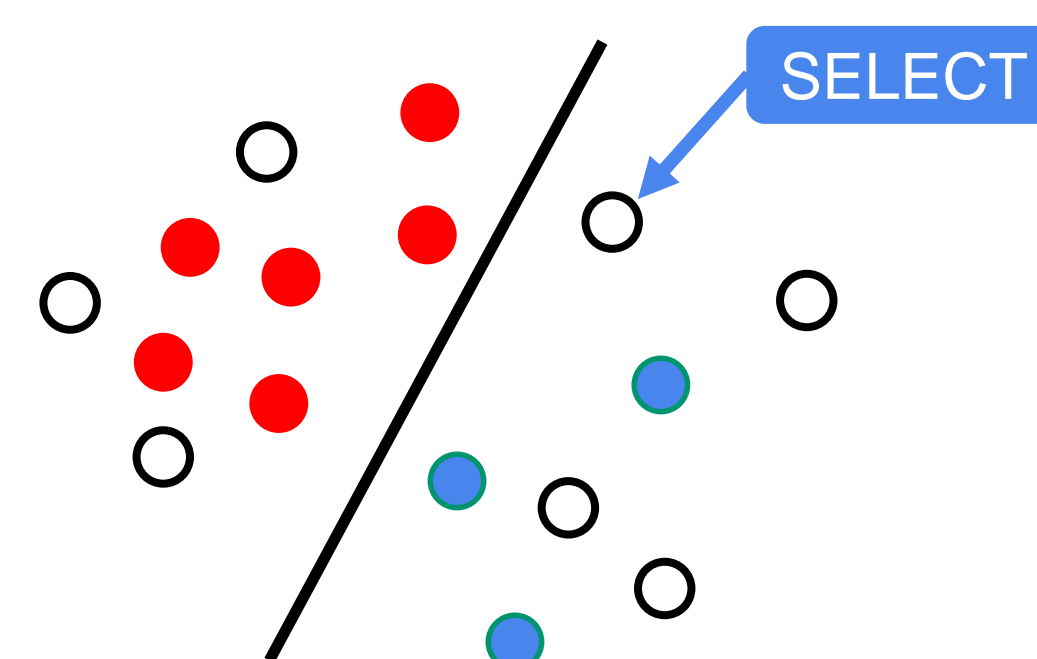
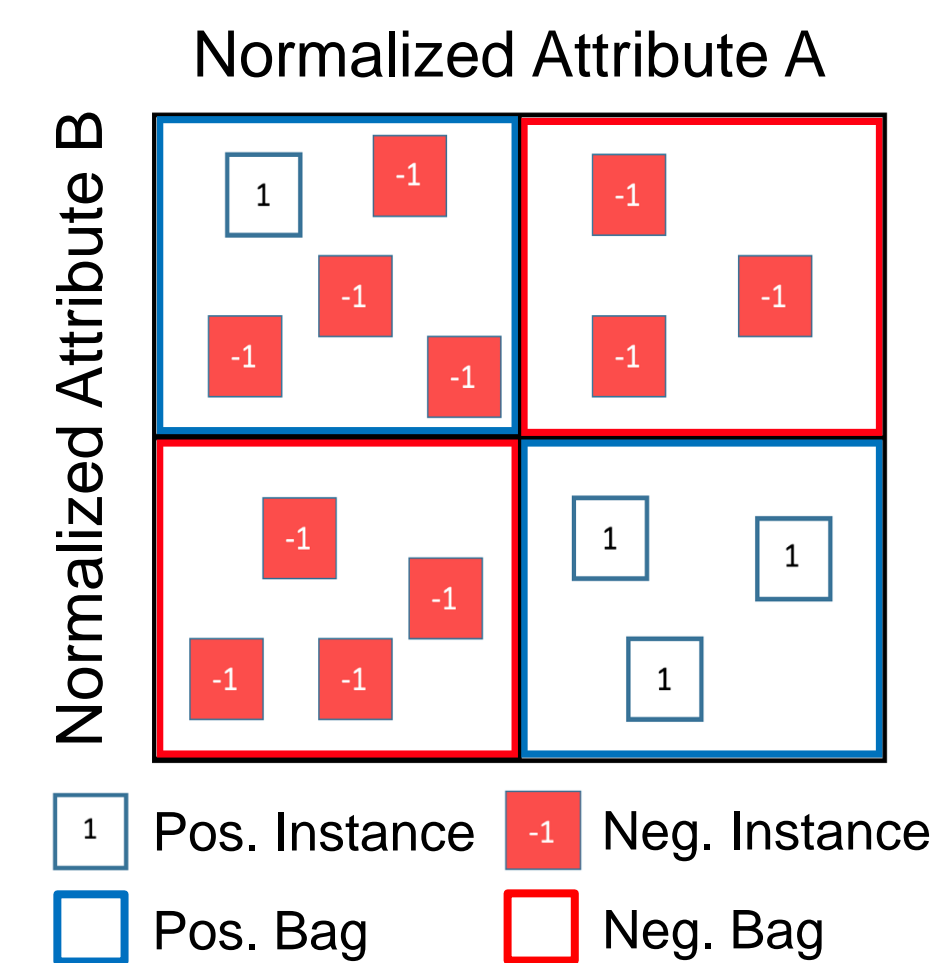
- ❖ QFE is an Automatic Data Exploration approach
- ❖ Relies solely on simple feedbacks provided by users
- ❖ Attractive choice for non-expert database users

The REQUEST Framework



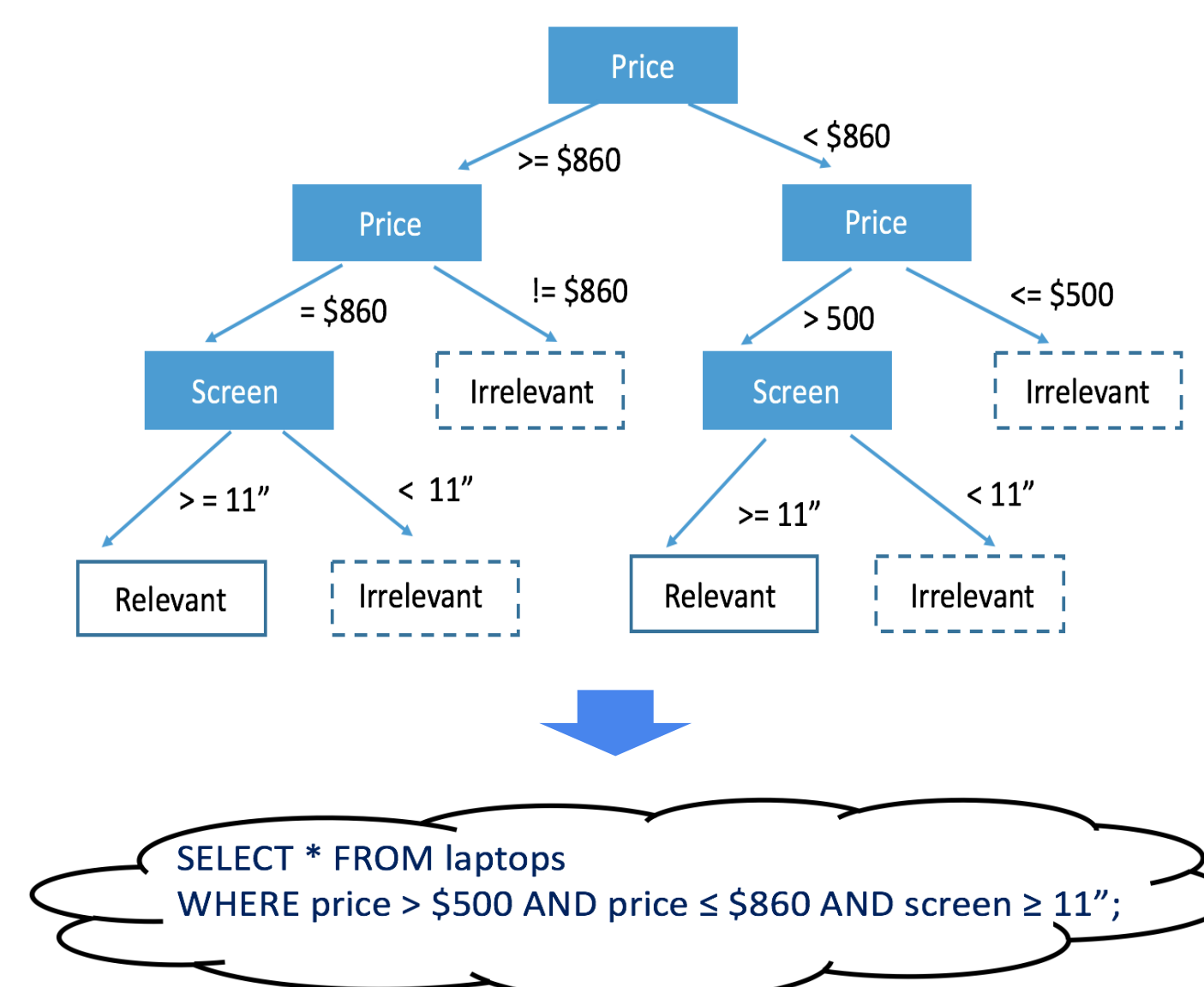
- ❖ Objectives:
 - Minimize human effort
 - Short response time
 - Auto query formation

- ❖ Data Reduction using User-Driven Pruning (UDP)
 - Objects are grouped into a set of bags
 - Feedback given on a bag of objects
 - Reduce search space by pruning irrelevant bags



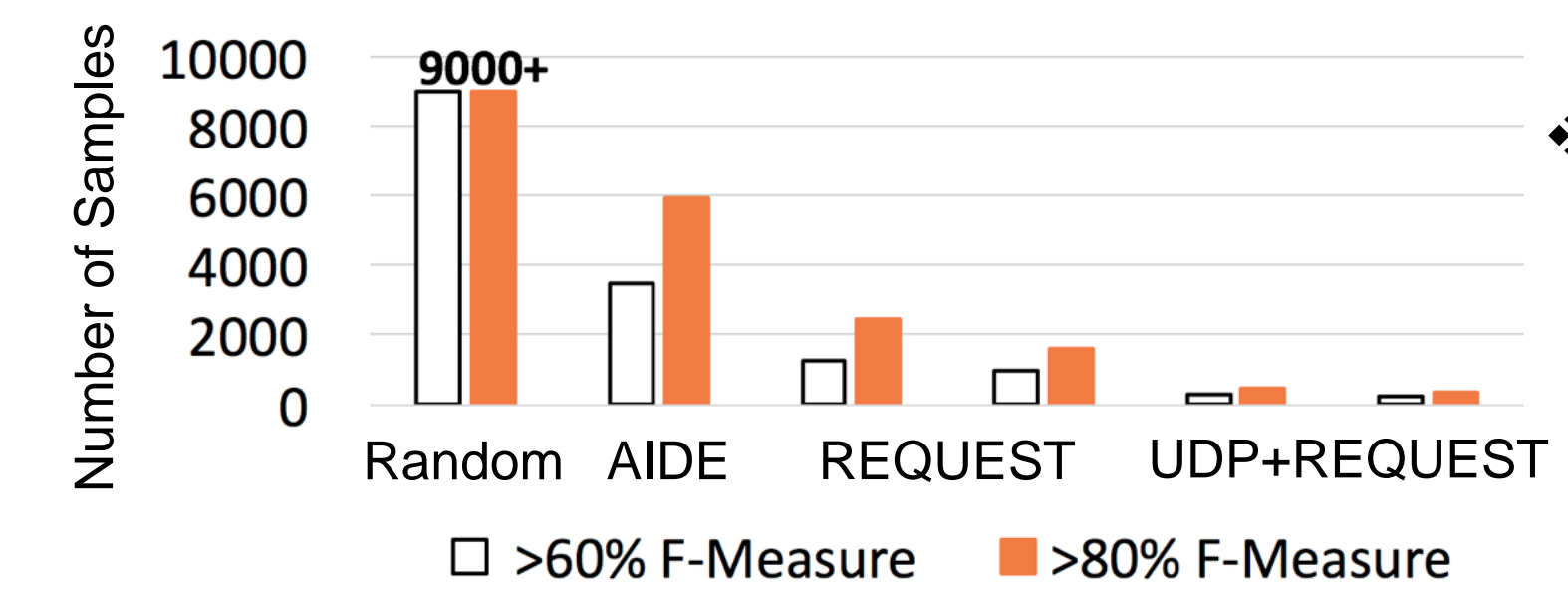
- ❖ Query Selection using Uncertainty Sampling:
 - Chooses the data point which is most uncertain to the current classification model as the next training sample

- ❖ Query Formation using Decision Tree Classifier:
 - Formulate the range selection query that select all the objects of interest

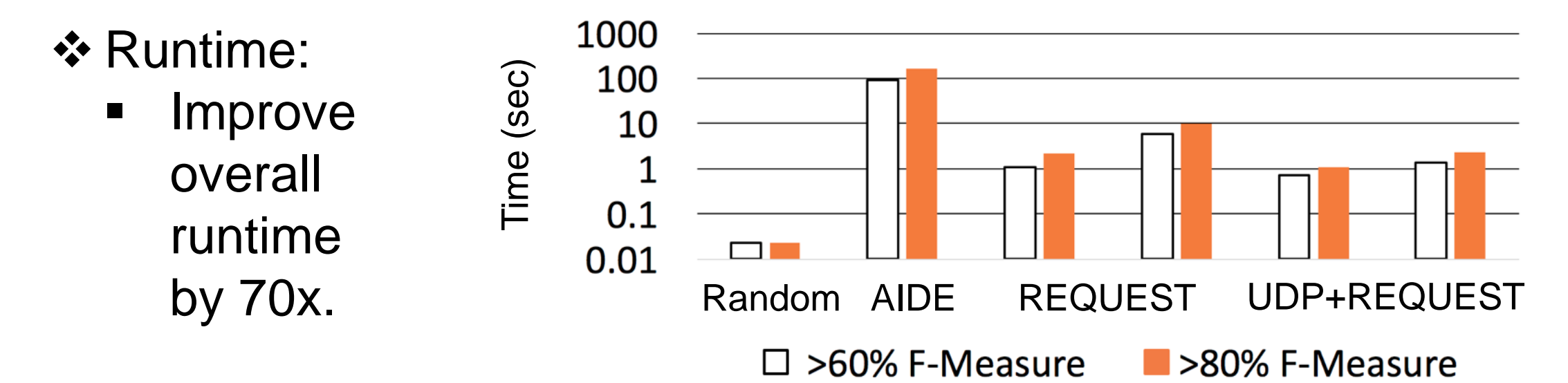


```
SELECT * FROM laptops
WHERE price > $500 AND price <= $860 AND screen >= 11";
```

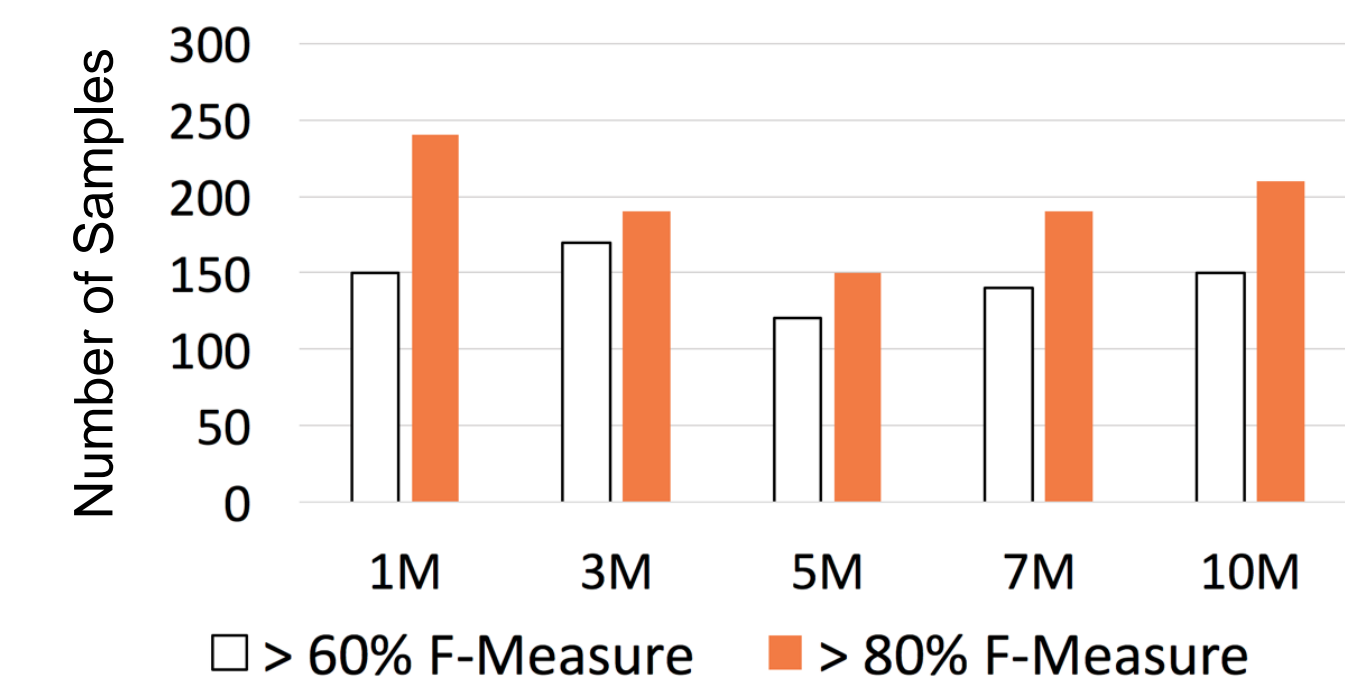
Experiments and Results



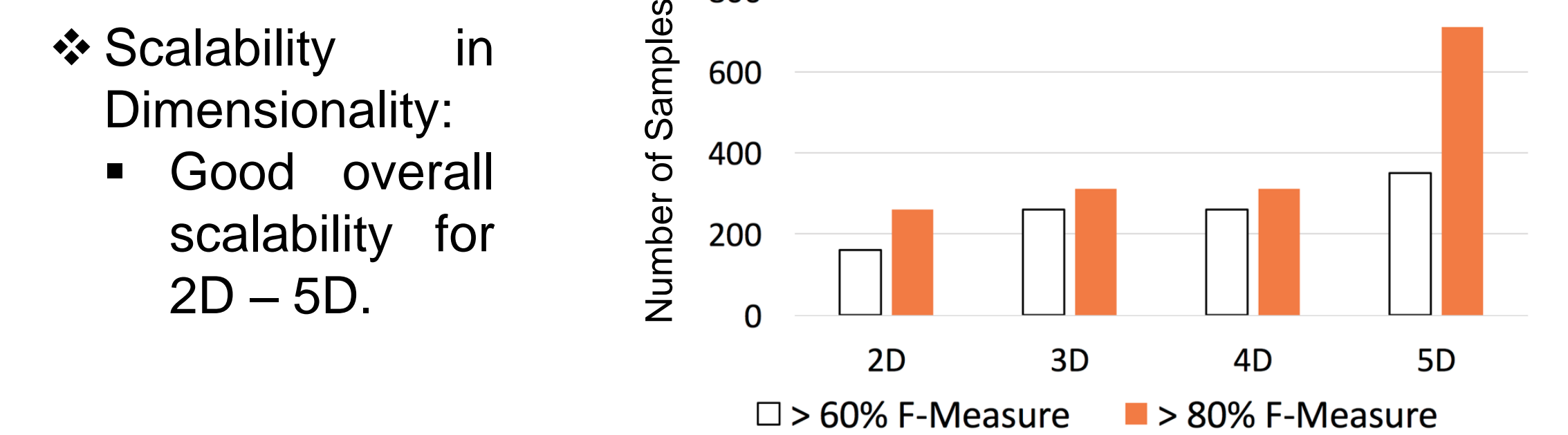
- ❖ Accuracy:
 - Reduce human effort by about 92%.



- ❖ Runtime:
 - Improve overall runtime by 70x.



- ❖ Scalability in Dataset Sizes
 - Good overall scalability with respect to different data sizes.



- ❖ Scalability in Dimensionality:
 - Good overall scalability for 2D - 5D.

Conclusion

- ❖ REQUEST is a novel data exploration framework.
- ❖ Interactive and scalable data exploration.
- ❖ Reduce human effort by 92%.