



IDLE: Interactive Descriptions for Low-dimensional Embedding

Aindrila Basak¹, Mona Nashaat¹, James Miller¹, Shaikh Quader²

¹University of Alberta, Edmonton, Canada, ²IBM Canada, Toronto, Canada

Research Objectives

Interpret Black-Box Models

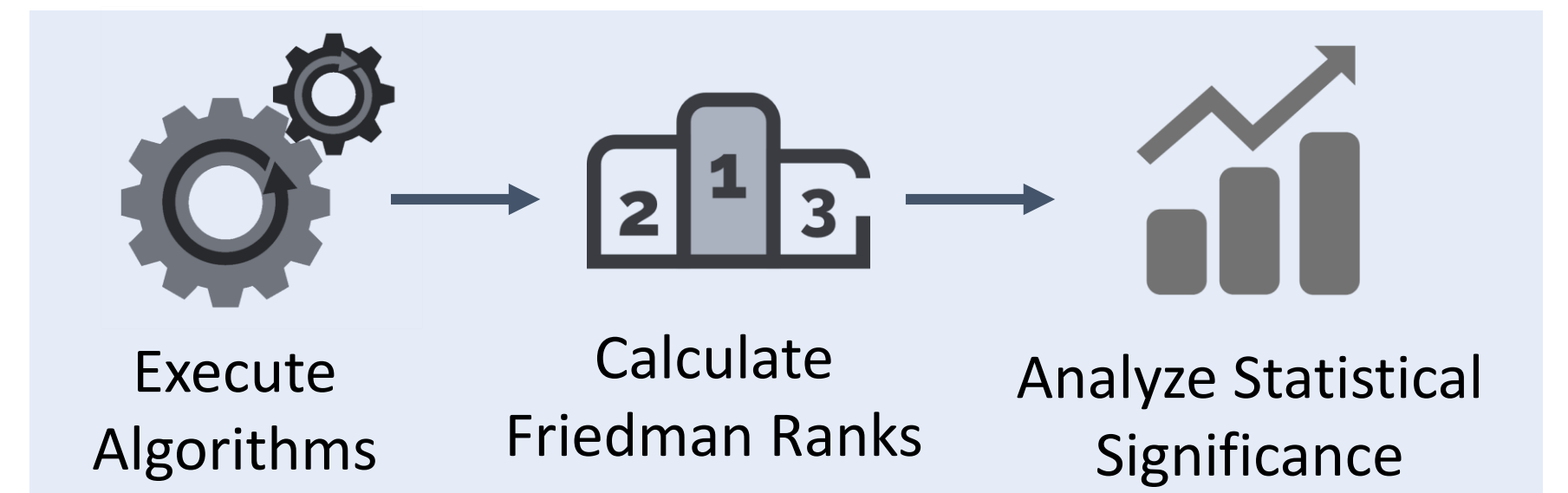
- Most big data analysis algorithms are black-box.
- Users do not use what they don't understand or trust.
- Leading to the loss of millions of dollars invested in building the models.

Engage Users in Decision Making

- Necessity for Human-in-the-loop.
- Need to incorporate user expertise in decisions.
- Need for proactive guidance with complex data analysis.

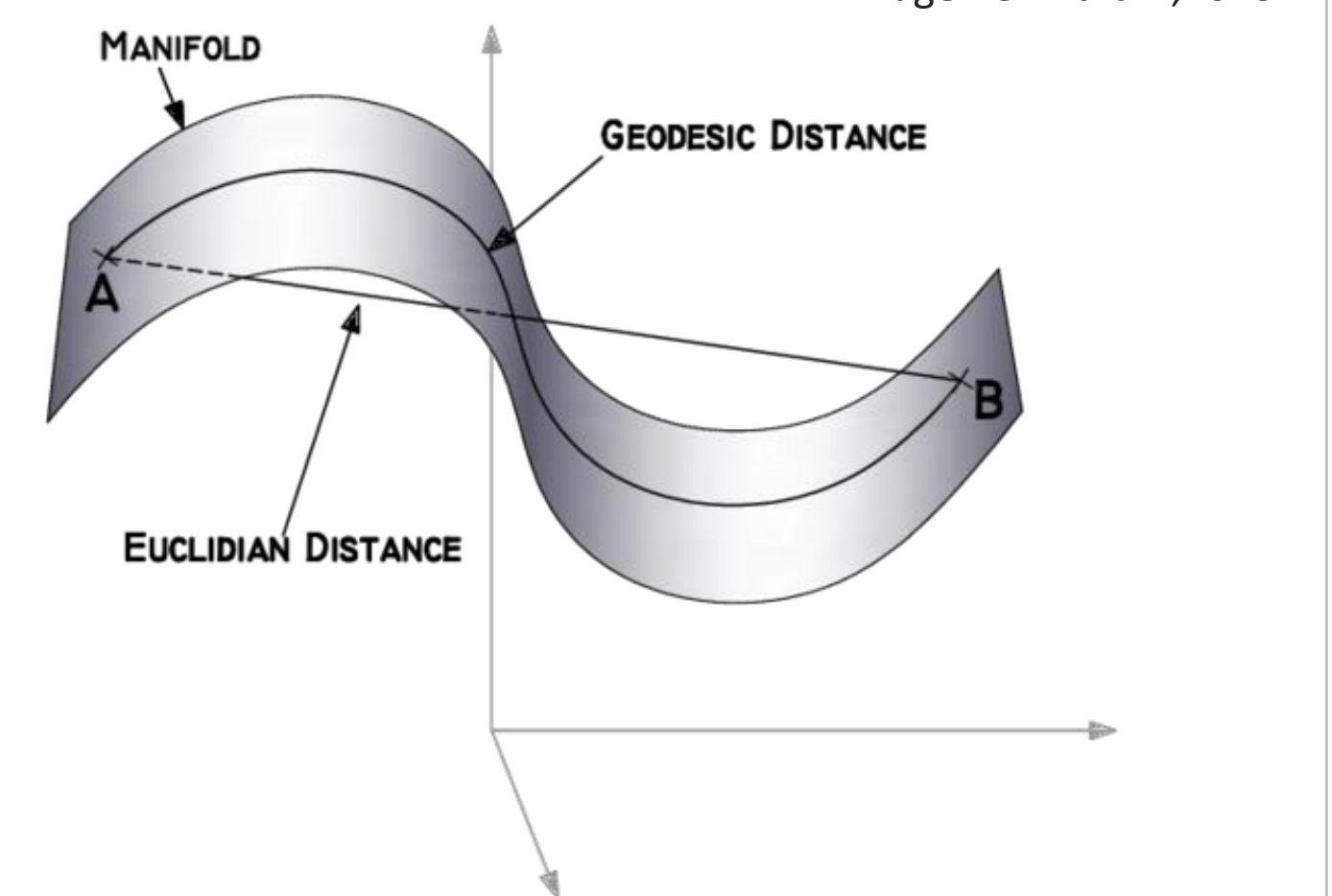
Experimental Methodology

- Compare 15 Dimensionality Reduction Algorithms
- For 7 Evaluation Metrics
- Over 30 Real-world Datasets
- Evaluate with 6 Statistical Significance Tests



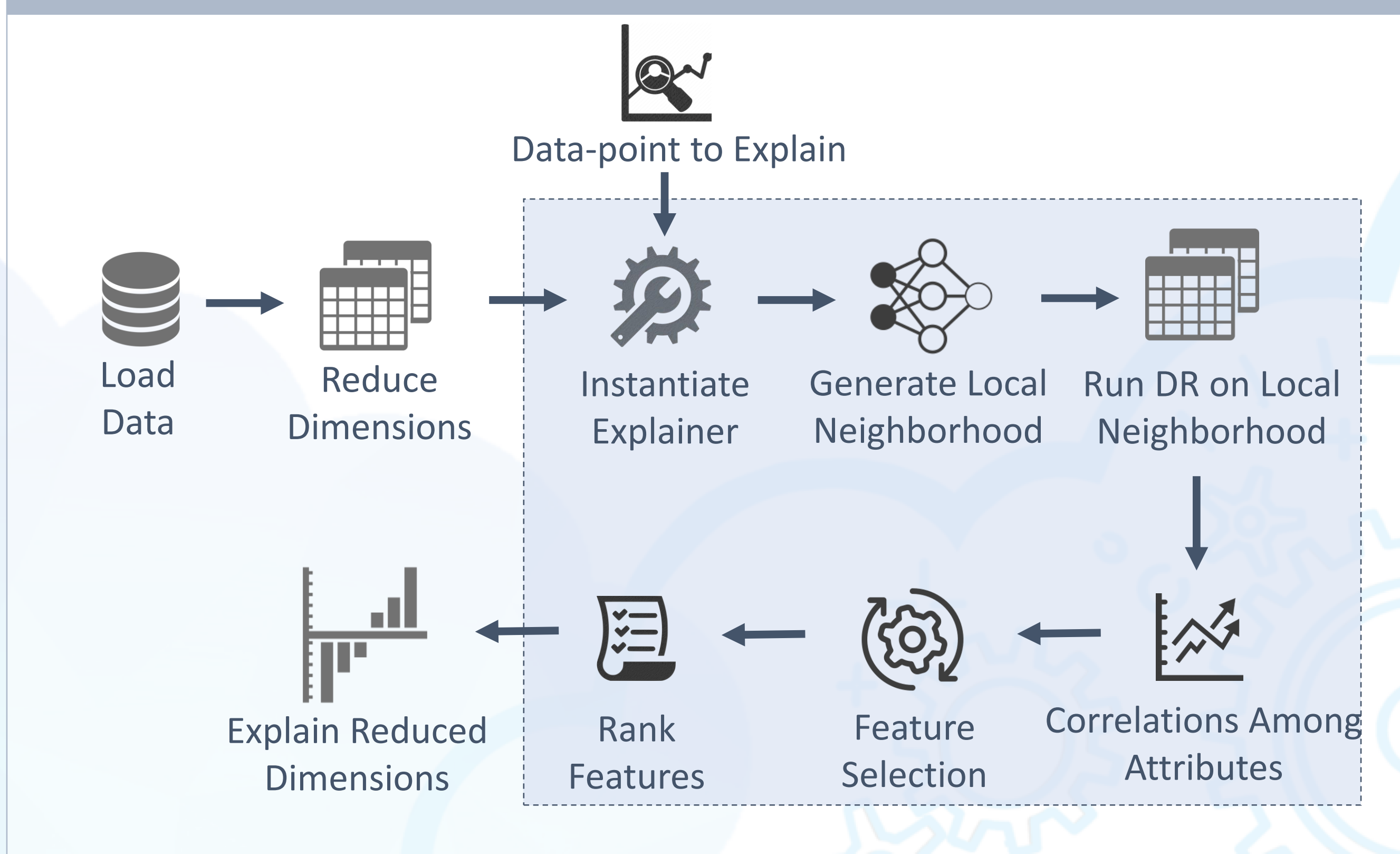
Experimental Outcome

Evaluation Metric	Best Performers	Mediocre Performers	Worst Performers
ML Accuracy	KernelPCA, PCA	Fit-SNE, LEM	LTSA, HLLC
Execution Time	PCA, Isomap	openTSNE, LTSA	MDS, LEM
Local Structure	MDS, openTSNE	Fit-SNE, UMAP	LLE, Isomap
Global Structure	MDS, KernelPCA	LEM, HLLC	Trimap, t-SNE
Outlier Effects	LTSA, Isomap	t-SNE, openTSNE	LLE, MLLC
Duplicate Effects	t-SNE, Trimap	HLLC, LEM	MDS, KernelPCA
Partial Records	PCA, KernelPCA	UMAP, Trimap	Fit-SNE, t-SNE



Preservation of Global Structure: Measure of pair-wise geodesic distances among data-points in the high-dimensional manifold.

IDLE - Process Flow



Results

