SEMDIFF: Explaining Semantic Changes in Evolving Databases

Shiyi He, Anna Fariha



How did my data change? Why?

name	gen	edu	exp	salary	bonus
Anne	F	PhD	2	\$230,000	\$23,000
Bob	M	PhD	3	\$250,000	\$25,000
Amber	F	MS	5	\$160,000	\$16,000
Allen	M	MS	1	\$130,000	\$13,000
Cathy	F	BS	2	\$110,000	\$11,000
Tom	M	MS	4	\$150,000	\$15,000
James	M	BS	3	\$120,000	\$12,000
Lucy	F	MS	4	\$150,000	\$15,000
Frank	M	PhD	1	\$210,000	\$21,000

Why is the **bonus** different

I wonder why some people

received a bonus increase,

such as Anne got a \$2150

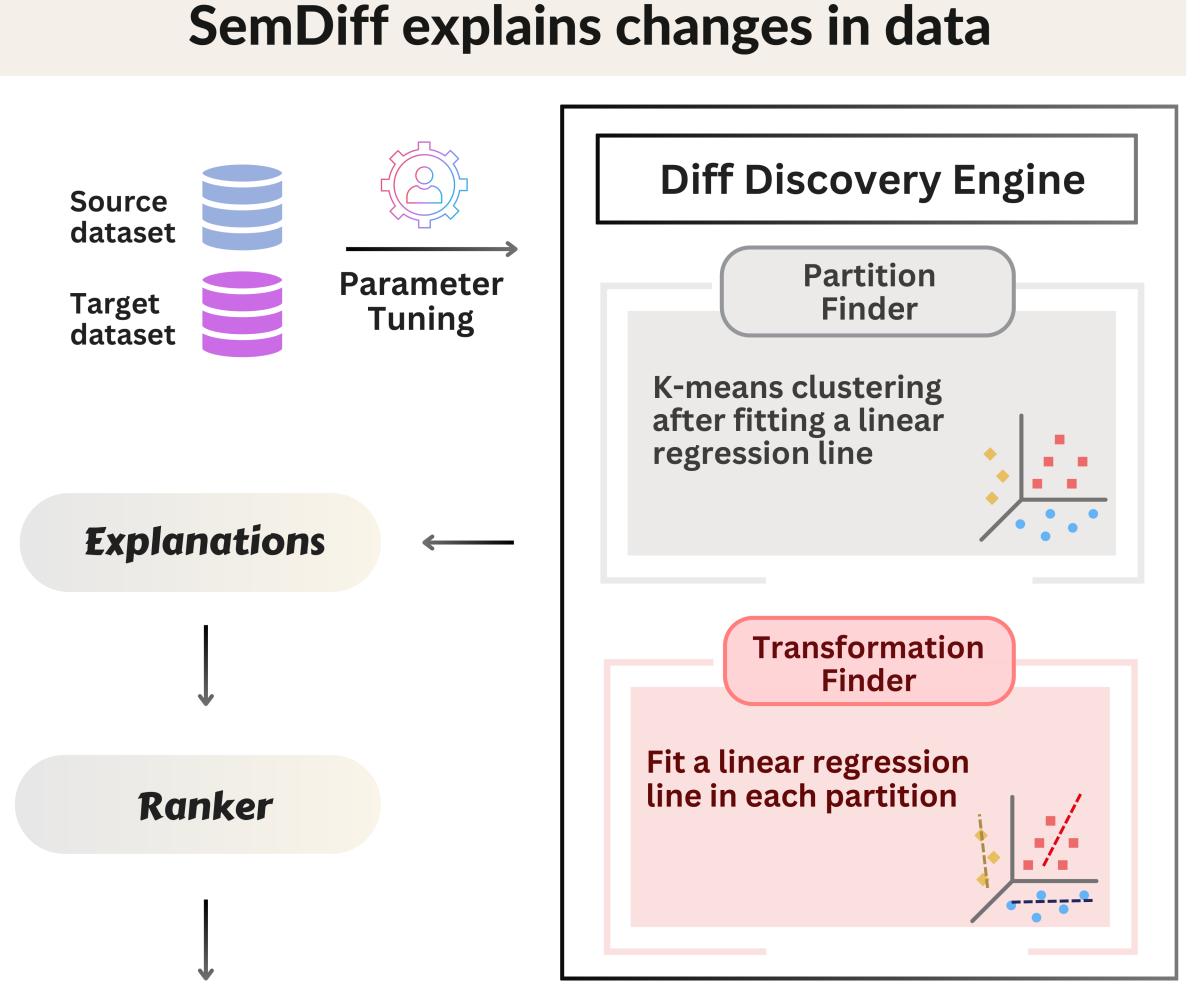
increase, while others like

Cathy didn't get any!

in 2017? What happened?

name	gen	edu	exp	salary	bonus
Anne	F	PhD	3	\$230,000	\$25,150
Bob	M	PhD	4	\$250,000	\$27,250
Amber	F	MS	6	\$160,000	\$17,440
Allen	M	MS	2	\$130,000	\$13,790
Cathy	F	BS	3	\$110,000	\$11,000
Tom	M	MS	5	\$150,000	\$16,400
James	M	BS	4	\$120,000	\$12,000
Lucy	F	MS	5	\$150,000	\$16,400
Frank	M	PhD	2	\$210,000	\$23,050

2016 data 2017 data





SemDiff strikes the right balance between accuracy and interpretability to generate the best explanation

Explanation Condition → Transformation





