Paper: <https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=6889613&tag=1>

Why feature selection?

* To reduce data dimensionality
* speeding up a machine learning or data mining algorithm
* improving learning accuracy
* Enhancing model comprehensibility
* better data visualization
* reduction of measurement and storage requirements

Why differentially private feature selection?

* Effectively mine sensitive data

What does the paper discuss?

* Local learning and differential privacy to achieve feature selection
* Few experiments on benchmark datasets

How does a feature selection algorithm look like?

* Two aspects to it: 1. Search strategy 2. Evaluation criterion
* Categorised into 3 models: 1. Filter model 2. Wrapper model 3. Embedded model
* Read more: <https://sebastianraschka.com/faq/docs/feature_sele_categories.html>
* Paper focuses on **feature weighting algorithm** and it belongs to **filter model**

How many ways can you preserve privacy in?

* 4 ways
* Input perturbation
* Transformed data release
* Query auditing & query answer perturbation
* Access control
* Paper talks about **query answer perturbation** and uses **𝜀-differential privacy method**
* What’s that? Measure of quantifying privacy risk. *A statistical procedure satisfies*  *𝜀-differential privacy if changing a single datapoint does not shift output distribution by much.* Crudely, it promises that the *probability of a point belonging to a set is at least* ***eε*** *times that of a neighboring point belonging to the same set.*

What paper does?

* Proposes algorithm named **FWELL(Feature WEighting algorithm based on Local Learning)**
* Applies **sensitivity-based method**

Feature selection in the paper

* Feature weighting based on **Local Learning** - For a given point, it finds the nearest point in the same cluster (NHi) (clustering is based on the labels), and the nearest point in a different cluster (NMi). Then it calculates

* **Logistic loss with L2-regularizer** used for evaluation.

, and

* Stochastic GD is applied on *L2* loss function and the weights ***w*** are updated.

What improvisations can you suggest for this algorithm (scope of research)?

* This algorithm handles feature selection and making it differentially private, differently. Both are sort of, standalone operations. **Instead of performing feature selection first, we can make the model differentially private first.**
* Paper suggests one of many possible feature selection algorithms. **We can propose a standalone feature selection algorithm upon which the later step of DP can be wrapped.**
* This paper handles mostly classification problems. **We can extend our feature selection algorithm to tackle regression problems as well.**
* **We can improvise on the way noise is added to make model differentially private.**
* Manhattan distance is not a metric. Use some other metric.