



NTT Secure Platform Laboratories  
NTT セキュアプラットフォーム研究所

**第一分科会**  
**「パーソナルデータの利活用はどうあるべきか」**

**『匿名が役に立つケースの紹介～お勧め料理ができるまで～』**

---

**2013年11月23日**

**NTT セキュアプラットフォーム研究所 藤村明子**

- ・ 10年来の楽園だった匿名～ビジネスの有用性
- ・ 匿名が役に立つケースについて
  - 個票から統計処理をする・・・というところで議論が止まる。その先は？
  - 集計処理して販売したい。：でも何に使うつもりだったんだらうというのがよくわからない
  - 最終的な料理の姿によって、食材の調理の仕方が決まるのに(煮るか焼くか茹でるか)、それが決まらないと処理はできない。
  - 半生で売るのはアブナイ。けれども料理の姿が決まっているのだったら、それに向けて調理の仕方を煮るか焼くか茹でるか決められる。
    - ・ 料理が決まってないのに食材だけ売ろうとするのがこれまでの事例
- ・ 中立の立場で
- ・ 架空の事例で
  - オンライン書店のおすすめリスト
  - これを買った人はこんなのも買っています(いわゆる、お勧めリスト)の例
- ・ わかりやすく
  - 分野横断的

# 生の食材【特定個人の識別可能なデータ】

- ・オンライン書店で書籍を購入した顧客別の販売データについて
- ・それらのうち、以下5名の3つずつのレコードを例にする（↓架空の事例です）

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3		....
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名	
aaaaaa	藤村	2013/9/10	倍返しだ!	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す	....
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ	....
ccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟	....
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ!	....
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな	....
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
...	...	...	...	...	...	...	...	....

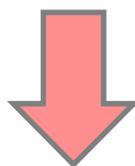
↑  
**特定個人の  
 識別性あり**

....	レコードn	
....	...	...
....	...	...
....	...	...
....	...	...
....	...	...
....	...	...
⋮	⋮	⋮
....	...	...

# 半生の食材【個人の識別可能なデータ】

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3	
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名
aaaaaa	藤村	2013/9/10	倍返しだ！	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ
ccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ！
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな



顧客の氏名を削除した

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3	
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名
aaaaaa	藤村	2013/9/10	倍返しだ！	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ
ccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ！
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな



**個人の  
識別性あり**

# NTT 食材を料理にあわせて加熱(煮る/焼く/茹でる)【レコードの組合せ】

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3	
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名
aaaaaa	藤村	2013/9/10	倍返しだ!	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ
ccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ!
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
...	...	...	...	...	...	...	...

顧客ID、氏名を削除、  
レコードの組合せIDを付与

同一顧客が30日以内に購入した  
レコードの組合せを抽出

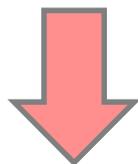
同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
1	aaaaaa	藤村	2013/9/xx	倍返しだ!	2013/9/xx	やられたらやり返す
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す

購入間隔が30日以内である  
ことさえ示せば、何でもよい



顧客IDを削除、  
レコードの組合せIDを付与



同一顧客が30日以内に購入した  
レコードの組合せを抽出

同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
1	aaaaaa	藤村	2013/9/xx	倍返しだ!	2013/9/xx	やられたらやり返す
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
3	bbbbbb	佐藤	2013/11/xx	統計に頼るな	2013/11/xx	万能ナイフはない
4	bbbbbb	佐藤	2013/11/xx	万能ナイフはない	2013/11/xx	座長はつらいよ
5	ccccc	森	2013/10/xx	孤軍奮闘	2013/11/xx	呉越同舟
6	dddddd	高木	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
7	dddddd	高木	2013/10/2	やられなくてもやり返す	2013/11/xx	100倍返しだ!
8	eeeeee	菊池	2013/11/xx	万能ナイフはない	2013/11/xx	統計に頼るな
⋮						
n	...	...	...	...	...	...

購入間隔が30日超のため外れたレコード

購入日	販売書籍名
2013/9/2	四面楚歌
購入日	販売書籍名
2013/4/1	ブロードウェイに寄り道

# 加熱中【集計処理をしている】

同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
1	aaaaaa	藤村	2013/9/xx	倍返しだ！	2013/9/xx	やられたらやり返す

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
3	bbbbbb	佐藤	2013/11/xx	統計に頼るな	2013/11/xx	万能ナイフはない

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
4	bbbbbb	佐藤	2013/11/xx	万能ナイフはない	2013/11/xx	座長はつらいよ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
5	cccccc	森	2013/10/xx	孤軍奮闘	2013/11/xx	呉越同舟

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
6	dddddd	高木	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す

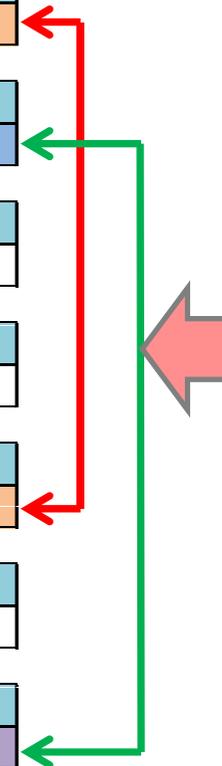
組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
7	dddddd	高木	2013/10/2	やられなくてもやり返す	2013/11/xx	100倍返しだ！

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
8	eeeeee	菊池	2013/11/xx	万能ナイフはない	2013/11/xx	統計に頼るな

⋮

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
n	...	...	...	...	...	...

同じ販売書籍名の組合せが  
複数回出現する組を抽出する



- \* 黄色とオレンジ
- \* 紫色と青色

統計の手法

同じ組合せが複数回出現しない組(マイナーな組)は削除する

# 加熱完了【複数回出現する組の抽出】

同じ販売書籍名の組合せが複数回出現する組を抽出する



同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
3	bbbbbb	佐藤	2013/11/xx	統計に頼るな	2013/11/xx	万能ナイフはない

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
6	ddddd	高木	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
8	eeeeee	菊池	2013/11/xx	万能ナイフはない	2013/11/xx	統計に頼るな

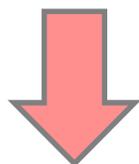
⋮

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
n	...	...	...	...	...	...



使わない属性はすべて削除

# お皿にのせる【目的とするデータ】



そして、販売書籍名のみを抽出

この本を買った人はこんな本も買っていますリスト 1

販売書籍名
やられたらやり返す
やられなくてもやり返す

黄色を買ってる人は  
オレンジも買ってる傾向がある

この本を買った人はこんな本も買っていますリスト 2

販売書籍名
万能ナイフはない
統計に頼るな

青色を買ってる人は  
紫色も買ってる傾向がある

この本を買った人はこんな本も買っていますリスト n

販売書籍名
...
...
...
⋮
...

・こうしていわゆる「お勧めリスト」になった

・今回のケースは2つだけだが実際はもっと多い数。  
多いほど相関関係があると推測できる

・お勧めリストでは、多数の人の購買情報が必要とされるから、稀少データは不要

・あらかじめ「何に使うのか」決まっていれば、このように処理もおのずと決まってくるはず

・匿名への(匿名性と匿名化)

特定個人の  
識別可能な  
データ

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3		...	レコードn	
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名		...	...
aaaaaa	藤村	2013/9/10	倍返しだ!	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す	...	...	...
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ	...	...	...
cccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟	...	...	...
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ!	...	...	...
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな	...	...	...
⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮	⋮
...	...	...	...	...	...	...	...	...	...	...

個人の  
識別可能な  
データ

顧客別販売データ

顧客ID	氏名	レコード1		レコード2		レコード3	
		購入日	販売書籍名	購入日	販売書籍名	購入日	販売書籍名
aaaaaa	藤村	2013/9/10	倍返しだ!	2013/9/30	やられたらやり返す	2013/10/10	やられなくてもやり返す
bbbbbb	佐藤	2013/11/1	統計に頼るな	2013/11/8	万能ナイフはない	2013/11/23	座長はつらいよ
cccccc	森	2013/9/2	四面楚歌	2013/10/29	孤軍奮闘	2013/11/23	呉越同舟
dddddd	高木	2013/9/27	やられたらやり返す	2013/10/2	やられなくてもやり返す	2013/11/16	100倍返しだ!
eeeeee	菊池	2013/4/1	ブロードウェイに寄り道	2013/11/1	万能ナイフはない	2013/11/8	統計に頼るな

同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
1	aaaaaa	藤村	2013/9/xx	倍返しだ!	2013/9/xx	やられたらやり返す
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
3	bbbbbb	佐藤	2013/11/xx	統計に頼るな	2013/11/xx	万能ナイフはない
4	bbbbbb	佐藤	2013/11/xx	万能ナイフはない	2013/11/xx	座長はつらいよ
5	cccccc	森	2013/10/xx	孤軍奮闘	2013/11/xx	呉越同舟
6	dddddd	高木	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
7	dddddd	高木	2013/10/2	やられなくてもやり返す	2013/11/xx	100倍返しだ!
8	eeeeee	菊池	2013/11/xx	万能ナイフはない	2013/11/xx	統計に頼るな
⋮	⋮	⋮	⋮	⋮	⋮	⋮
n	...	...	...	...	...	...

同一顧客が30日以内に購入した書籍の組合せデータ

組ID	顧客ID	氏名	購入日	販売書籍名	購入日	販売書籍名
2	aaaaaa	藤村	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
3	bbbbbb	佐藤	2013/11/xx	統計に頼るな	2013/11/xx	万能ナイフはない
6	dddddd	高木	2013/9/xx	やられたらやり返す	2013/10/xx	やられなくてもやり返す
8	eeeeee	菊池	2013/11/xx	万能ナイフはない	2013/11/xx	統計に頼るな

目的とするデータ

\* どの段階で販売するのか

販売書籍名
やられたらやり返す
やられなくてもやり返す
販売書籍名
万能ナイフはない
統計に頼るな