PUVIT PRACHARKTAM 6031830321

HOW DOES A SUPERMARKET USE DATA ANALYTICS TO PREDICT PREGNANCIES

Firstly the retail must have club card or something that can match customer id with product they buy similar to All Member in 7-11. Then, the customer have to collecting the data which customer is pregnant by providing promotion or something, in order we need the material to training and testing our model in the future. Next, cleaning data is very necessary, from parsing raw data to .csv file. And filtering duplicate data and outliers data for example, men cannot be pregnant numeric data with very little possibility (people 150 years old). So, we can decide later whether we erase that data or edit that data by using statistical methods.

Now we got customer data already, pregnant women have some pattern of buying the same product in each period of time. We can use this pattern to create pregnancy prediction model like this *Pregnancy Score = f(item1, item2, item3, … , item25).* Pregnancy Score will run between 0-1 (or equivalent to 0-100%), predicting how much probability a female customer is pregnant. So we got the training set as below.

Customer_id	Lotion	Vitamin	Calcium	Cocoa-Butter	Cotton	…	is_pragnant
1	1	1	1	1	0		1
2	0	1	1	0	0		0
3	1	1	0	0	1		1
4	0	0	0	1	0		0

Once we has information about transactions for pregnant clients, scoring / prediction can be done easily. basic classification algorithms such as **Logistic Regression** or **Decision Tree** etc.