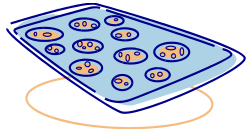


Data Mining / Partitioning Example

Scenario: You own a small bakery and you want to bake and sell the best chocolate chip cookies in town. You suspect that putting more chocolate chips in your cookies may increase the number of customers who like your cookies. You decide to use a data mining experiment to find out if you are correct.

Training Data Set



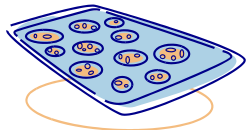
You bake a *first* batch of cookies with varying amounts of chocolate chips



You have your customers do a "taste test" and indicate if they "like" or "don't like" each cookie that they taste

You detect a pattern: Cookies with 10 or more chips tend to get more "likes" from customers

Validation Data Set



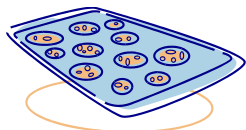
You bake a *second* batch of cookies to validate the pattern. Half the batch has <10 chocolate chips, the other half has ≥ 10 chips.



Again, you have your customers do a "taste test" and indicate if they "like" or "don't like" each cookie that they taste

You discover that 10 out of 12 (83%) customers "liked" cookies with 10 or more chips

Test Data Set



You bake a *third* batch of cookies to sell in your bakery. All of the cookies have 10 or more chips.



You sell your new cookies in your bakery and hope that 83% of customers will like them!