**Techpoint AI Model Overview**

a) Picking the Right Tool (Model Selection):

* **Look at the Data**: We have mostly category-type info, like different food orders.
* **Best Tools for the Job**:
  + **Decision Trees**: Like a flowchart – asks questions about data and makes decisions.
  + **Random Forest**: Many decision trees combined – better decisions by working together.
  + **Gradient Boosting**: Corrects its own mistakes. Strong performer but needs more computer power.
  + **Logistic Regression**: Simple tool, checks how factors, like the university someone studies at, might influence their order.
* **Others We Didn’t Pick**:
  + **SVM & Neural Networks**: Great for complex problems but might be too much for this.
  + **KNN**: Like asking neighbors for advice. Not great for big data.
* **How We'll Judge Success**:
  + We’ll see how often the model’s guesses are right, and also how well it spots true positives (like correctly guessing a specific food order).

b) Training and Testing:

* **Splitting the Data**: We'll set aside some data to train our model and some to test it.
* **Training**: The model learns from known data.
* **Testing**: We see how good the model is with unknown data.
* **Tweaking the Model**: We’ll adjust some settings to get better results.

c) Preparing the Data:

* **Cleaning Up**: We remove any errors or gaps in the data.
* **Adding Insight (Feature Engineering)**: Think of patterns. Do students from a culinary college order different foods?
* **Changing Data Types**: The model needs numbers. So, we’ll turn categories (like university names) into numbers.
* **Balancing Data Sizes (Scaling)**: Makes sure no piece of info has too much say in our model's decisions.
* **Handling Sparse Data**: Some universities, like Purdue, have few orders. We might group such universities as "Others" since one or two orders don’t really tell us the full story about all Purdue students.