NLP Questions

1. You just trained a model to classify soft drinks according to their flavor-profile. The results are below. What has the model learned about the correspondence between the item name and the classification? Can you find any problems with what the model has learned? How would you address this problem?

|  |  |
| --- | --- |
| **item\_name** | **flavor\_profile** |
| 7up can 355ml | lemon-line |
| pepsi bottle 14oz | cola |
| 7up bottle 500ml | lemon-line |
| 7up can 500ml | lemon-line |
| pepsi bottle 20oz | cola |
| coke bottle 14oz | cola |
| pepsi can 20oz | cola |
| coke bottle 20oz | cola |
| sprite bottle 500ml | lemon-line |
| sprite can 500ml | lemon-line |
| sprite can 355ml | lemon-line |
| coke can 20oz | cola |
| sprite bottle 355ml | lemon-line |
| 7up bottle 355ml | lemon-line |
| pepsi can 14oz | cola |
| coke can 14oz | cola |

1. You’ve been given a product list of ~16,000 carbonated soft drinks (see carbonated\_soft\_drinks.csv) and you’re asked to build a model that can extract the product’s brand (e.g., Coke, Pepsi Sunkist), when available. What type of model would you use? If you choose a supervised-learning model, where will you get your training data?