Warehouse Storage Optimization - Report 2

Dhruvil Dave, Dhatri Kapuriya, Harvish Jariwala, Nisarg Thoriya
School of Engineering and Applied Sciences
Ahmedabad University
Ahmedabad, Gujarat, India - 380009
dhruvil.d@ahduni.edu.in (AU1841003)
dhatri.k@ahduni.edu.in (AU1841129)
harvish.j@ahduni.edu.in (AU1841050)
nisarg.t@ahduni.edu.in (AU1841142)

***Abstract*—This is the second progress report of our group *Gopher - Group 4* for Machine Learning (CSE523) course project.**

***Index Terms*—classification, data pre-processing, clustering, graphical models**

I. Introduction

For our project, we decided to use the Amazon Bin Images Dataset.

This is originally a Computer Vision dataset. The Amazon Bin Image Dataset contains over 530,000 images and metadata from bins of a pod in an operating Amazon Fulfillment Center.

II. Task Performed And Outcomes

A. Exploratory Data Analysis

The data consists of seven columns each explaining the product features. They are height, length, asin(Amazon Standard Identification Number), qnty, weight, width and bin. Our main task for this week was to find out the output variables and input features. The data is distributed in a way that we have one bin column which states the bin number. The product belonging to any bin falls in that bin number respectively.
For optimising the storage problem we need to assume that size of all the bins are same. In the dataset we are not given the dimensions of the bin. Our predictor function will have product dimensions *i.e.* height, weight, width, weight and asin as input features. The output would return the bin number in which is must be placed.

B. Brief insight of modelling

Suppose, we have an empty storage and the ware house is new. We would train our model from the initial stage such that when any new product is to be placed the algorithm would look for two cases: Empty bins and Half filled or entirely filled bins. If there are bins who are half filled then model would proceed to look for some storage space left in any bin and would check for best possible bin to be placed into.

C. Progress

The foremost step would be to find the bin size in which products are stored. Hence, we found the volume occupied by products inside each bin.

```
import pandas as pd
import sqlite3

con = sqlite3.connect("clean_data.sqlite")
df = pd.read_sql_query("SELECT * from inv", con)
df['volume'] = df['height'] * df['length'] * df['
    width'] * df['qnty']
```

	height	length	asin	qnty	weight	width	bin	volume
0	1.1	18.0	B018240DGG	3	0.600000	11.7	0	694.979998
1	0.9	8.9	1593859864	1	0.900000	6.0	1	48.060000
2	1.4	6.5	B0178Y7KVM	5	0.022046	5.0	1	227.499999
3	3.2	9.9	B000052Z9F	1	2.250000	3.4	2	107.712000
4	2.3	7.4	B000HM5RPO	1	0.700000	5.6	2	95.312000
...	...	...	...	...	...	...	...	...
1384704	3.4	8.6	B00EIOYICA	1	2.900000	3.6	536433	105.264000
1384705	2.2	3.4	B016MD0TBI	4	0.022046	3.3	536433	98.736000
1384706	1.0	11.2	B017NEZ73A	1	0.620000	8.0	536433	89.600000
1384707	5.0	6.9	B019HI4CRC	2	0.750000	5.4	536433	372.599999
1384708	2.1	6.3	B01D8030L4	1	0.750000	6.2	536433	82.026000

Fig. 1. The dataset we obtained from amazon with added volume

Now, we grouped the bin and calculated the maximum volume stored by a bin *i.e.* we added the volumes of all products in each bin and found MAX out of it.

```
df[['bin', 'volume']].groupby('bin').sum().
    sort_values(by=['volume'],ascending=False)
```

	volume
bin	
122861	31724.243211

This implies that size of bins must be greater than or equal to the MAX volume. For the data we have, we will take it to be equal to the volume.

III. Tasks for Upcoming Week

The main tasks to be performed in the upcoming week are:

1) Exploratory Data Analysis (EDA)
2) Feature Engineering