

Armando Tejeda
Homework 14

Question 20.1

Describe analytics models that could be used to help the company monetize their data: How could the company use these data sets to generate value, and what analytics models might they need to do it?

There are lots of good answers, and I want you to think about two types – at least one of your answers should be based on just one data set, the one they’ve collected internally on customer browsing patterns on the web site; and at least one of your other answers should be based on combining more than one of the data sets.

Think about the problem and your approach. Then talk about it with other learners and share and combine your ideas. And then, put your approaches up on the discussion forum, and give feedback and suggestions to each other.

You can use the {given, use, to} format to guide the discussions: Given {data}, use {model} to {result}.

Here are the three data sets to consider:

DATA SET #1 (purchased from an alumni magazine publisher)

- first name
- last name
- college or university attended
- year of graduation
- major or majors
- marital status
- number of children
- current city
- email domain
- financial net worth
- binary variables (one for each interest in the publisher’s long list of various sports, activities, hobbies, games, etc.) showing whether each one was or wasn’t listed by each person

DATA SET #2 (purchased from a credit bureau)

- first name
- middle name
- last name
- marital status
- sex
- year of birth
- current city

- whether they ever owned real estate
- email domain
- list of monthly payment status over the last five years for credit cards, mortgages, rent, utility bills, etc. – for each month and each payment:
 - what type of payment it was – for credit cards, it would say “Visa”, “American Express”, etc., not just “credit card”
 - how much was owed
 - how much was paid
 - whether the person was considered to be in default

DATA SET #3 (collected by the company using web site tracking code)

- title
- first name
- middle initial
- last name
- credit card type
- credit card number
- list of products purchased in the past, with date of purchase and ship-to address
- which web pages the person looked at
- how long the person spent on each page
- what the person clicked on each page
- estimate of how long the user’s eyes spent on each page viewed (for customers where the software was able to take over the device’s camera)

Solution:

Model 1: This model will only use data available in dataset 3. Given how limited and basic this database is, I believe the best approach would be to rank the categories of purchases products over a timeframe and use that information to get some sort of a quick ROI.

Model 2: this model would use at least 2 of the databases available. This model will more accurately target audiences based on their interests and other factors.

Fore work:

Matching Data – before datasets can be used together all databases need to be integrated. This can be done manually using a variety of computational tools that will also make sure duplicates are removed.

Model 1:

Only dataset #3 – GIVEN previous search data and click activity data USE a linear or Tree Regression Model or Random Forest model TO create a score for all the items in inventory from the model to suggest the most interesting items the user is most likely to buy, increasing profit.

This is a heuristic approach for a quick ROI that relies on linearity of the expected outcome in relation to the variables.

Model 2:

Multiple datasets – Given integrated datasets 1,2 and 3, use tree model to categorize users into buckets that can be given a value. If enough data points are provided, give the large number of factors available, our tree model could render an accurate model. This classification of website users can be used in conjunction to model 1, which is primarily ranking the items in inventory, to target customer's interests.

This could also be used in conjunction with a multi-armed bandit model – a model very popular in marketing that continues to fetch information on customer preferences whilst moving closer to an optimal solution.