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)

Mihir Tulpule, Ali Lakdawala, Anu Rana, Kunle Lawal
ISYE 6501 - Introduction to Analytics Modeling
Professor Sokol
Homework 14

Solution Framework:

We began tackling this problem by first outlining the basic steps for our analysis, breaking down the problem into individual components and creating separate data/modeling techniques for each of the respective desired outcomes.

We began by asking which high-level questions we wanted to tackle and organizing them chronologically and determining what data might be needed to answer them:

1. The first step in this analysis would be to cluster the data. Combining data set 1, 2 and 3 we can use a k-means clustering, to create customer segments that can be individually targeted. Marketers frequently use clustering algorithms to determine how to bucket their audiences. In this situation running a clustering algorithm could show us which features are important or shared between different customers.

- Apart from k-means we can use other clustering methods such neural networks like autoencoders or a self organizing map.

- In essence due to the several unique dimensions of our dataset any clustering algorithm that is unsupervised and based on dimensionality reduction would be a good choice to derive insights and correlations from the data.

- This analysis would feed into the next steps in our solution framework.

- A drawback of using this approach is choosing the cluster centers and understanding what the cluster centers mean. We would need to tune our model and also have an understanding of what features it is using to create cluster centers.

2. Target products to customers based on interests and financial situation. Assuming we can successfully join datasets 1, 2 and 3 based on the first, last names and middle initial, we can use logistic regression, to predict the product customers are most likely to buy.

- This could be done using only 1 of the datasets versus all of them as well.

- Our analysis from step 1 could feed into making these models.

3. Create “ideal customer” profiles, look for others like these. From the analysis of 1 and 2 the company at this point should have been able to classify what their best customers are and their attributes are.

- Particularly in advertising and marketing once you've created your ideal customer profile you can use this to scale and find other customers who are similar to your ideal customer profile. All major advertisers, Facebook, Google, Pinterest, Adobe offer this tool to find new customers based on the profile of your current customers.

- This is a dual classification method. One classifying your own best customers you use a larger audience pool and classify those users as similar to your best customers.

4. Determine pricing for a customer based on financial situation.

- Given prior credit bureau information, net worth/demographic data from alumni magazine, purchase history/interests from retailer website use optimization to determine optimal pricing for each customer based on their given financial situation.

- A drawback of this approach is that if customers find out you're using this tactic, it might generate negative publicity for the company.

- This is becoming quite common though and heavily used by online clothing retailers and even amazon to price discriminate between customers based on their willingness to pay.

- This analysis could be extended to determine willingness to pay using Regression.

5. Rate colleges on value of degrees

- Given net worth of alumni from alumni magazine, credit ratings/personalized financial information from credit bureau for each individual alumna/alumnus use regression to assign rating to college based on predetermined scale.

- This information could be sold to major college ranking websites such as US News and Princeton Review which would be interested in tracking in what college alumni are doing after graduating to assign better ratings.

- An internal ranking could also be useful to determine which colleges this retailer should prioritize. They could build loyalty with students who have potential to convert to high net worth customers.

6. Using only dataset 3, we can build a model to predict the products customers will buy based on their web page viewing habits. Given dataset 3, use tree based models to classify which products customers will buy.

- This analysis can be supplemented by the analysis conducted in steps 1 and 2 of the solution framework.

- The viewing habits can be extended to dynamically updating the website based on where users are most likely to be looking on the website or the number of clicks they make before purchasing.

- The retailer could make the the path to purchase shorter for people who exhibit viewing habits with less attention to detail and more intention to purchase. While also adjusting the details shown to users based on the specific areas of the website they were focusing on.

7. Dynamic credit ratings, updated offers made for people with good alumni standing, credit rating and purchase behaviors.

- Given personalized credit information from credit bureau (past ratings, payment history, etc.), spending trends/purchase history from retailer website, occupation information from alumni magazine use regression to create dynamically changing credit ratings in real time based on updates from other three datasets as per availability.

- Although the US is still to implement a true dynamic credit rating system, this is already a norm in many countries around the world in Europe and China.

- Credit offers could be made to customers who show dependability of paying back. This flexibility would increase customer loyalty. These credit offers could also be in the form of a free trial and not necessarily monetary credit. Customers who are more likely to convert or purchase could be given these trial deals with the expectation that they're likely going to buy it anyway.

8. Provide their own store credit cards to high-worth customers or customers with good credit history.

- Given dataset 1 and 2 (and assuming we can again join them using the name information), use logistic regression, to predict the probability that they should get a store credit card.

- A new variable can be created based on the amount owed, which can represent whether a customer would get a store credit card or not. Then this new variable can be used as the response for our logistic regression.

- When we have new customers join/ makes a purchase we can use this model to predict whether we should immediately offer them this credit card or not based on the probability that they are high-worth customers.

9. Match underemployed people with companies hiring in area of interest.

- Given past personal credit history/financial information from credit bureau, previous occupation/interests and education information from alumni magazine, demographic information (e.g. area of residence) and past employment history from credit bureau use classification to match unemployed people with companies hiring nearby or in a historical area of interest.

- This information could be sold to Recruiting companies who are looking to fill roles.

10. Optimize pick up locations for customers of a certain product.

- Given dataset 3 (specifically product purchased, date of purchase and ship to address) along with existing pick up locations, use optimization, to assign the products to different pick up locations.

11. Estimate how long a customer would take to pick up a product, to free up storage space and for planning purposes.

- Given customer buying behavior data as well as demand data, a regression model could be used to estimate how long it takes for a customer to pick up their product. This leads to optimizing storage for delivery and pickup.

- Several large retailers such as Amazon, Walmart and Home Depot use these methods to efficiently pack lockers and not fill lockers for customers who are unlikely to pick up immediately.

12. Provide advertising to customers based on preferences and purchasing behaviors.

- Given past purchase history, most visited/purchased-from webpages, products bought together from retail website, income/personalized financial information from credit bureau, specific personal interests/hobbies from alumni magazine use design of experiments, A/B testing provide personalized and optimized advertising for each customer.

- A multi-armed bandit approach would work well to test different advertisements and continuously improve offers to customers.

13. Modify offers based on success rate or conversion rate of customers to the offer.

- Similar to 12, given customer online buying behavior data, the conversion rate of from product views to sales can be collected. A logistic regression model can be build to predict for whether a product will convert or not. For those products with high conversion rates, offers could be sent to customers viewing those products to improve it conversion rate. For products that do not have a high conversion rate, different kinds of offers could be sent to customer to improve the product conversion rates.

14. Optimize inventory levels. Given dataset 3, use time-series models (ARIMA), to obtain product purchasing trends.

- Next, given product purchasing trends, use simulation to test out different inventory systems and find the optimal one.

- Ideally probabilistic analysis of product purchasing patterns over time can be conducted which can be used to simulate the inventory over a period of time. Different inventory models can be compared and the system can be optimized.

15. Optimize search results by making website dynamic for individual customers based on preferences and past purchases

- Given customer buying behavior, a clustering algorithm could be used to group customers. And for customers in similar groups/clusters, a targeted search result product recommendation could be sent to those customers.

16. React and predict demand spikes.

- Given previous sales demand data, a time series model could be used to forecast demand for future time periods. This allows the retailer/store to be able to handle demand shortages.

17. Detect fraudulent transactions or retailer websites

- Could also be sold back to credit bureau, banks and credit card companies.

- Given past purchase history from retailer website, past credit/payment information from credit bureau, past interests/personal information from alumni magazine use logistic regression to determine the probability that a particular transaction is fraudulent or not.

<u>Summary Analysis:</u>

This homework scenario was initially difficult to tackle due to the sheer breadth of possibilities available from the three datasets and their resultant business applicability. We first divided up the potential use-cases into two categories: possible uses for the data from the three given original datasets, and possible uses for the company's own derived dataset (based on the three initial sets). Given that we were initially working with three datasets that could allow us to create detailed profiles of each individual customer (provided we could aggregate and match customers across datasets), we then utilized said data to tackle problems such as customer segmentation, fraud detection, dynamic updating of credit ratings, credit and promotion (e.g. store credit card) matching, and personalized pricing/advertisement targeting. When we thought about possible use-cases for the company's own dataset, we decided that this particular information lent itself to optimization and forecasting scenarios (contingent upon information also derived from the initial three datasets). As a result, we were able to use the company's dataset to modify logistics (e.g. inventory levels), manipulate the company website, provide personalized advertising, and modify offers to maximize profits. Although there are probably many more potential use-cases (e.g. using facial movements or predicted sentiment to construct a customer reaction analysis), we decided to only include use-cases that were specifically based on the data that we were given. Furthermore, we tried to identify use-cases that would be the most profitable for the company without jumping through unnecessary or costly hoops for relatively insignificant gains.