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:
  - o what type of payment it was – for credit cards, it would say "Visa", "American express", etc., not just "credit card"
  - o how much was owed
  - o how much was paid
  - o 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)

I am assuming that this is a large retailer. Here are some approaches that the company could monetize the datasets:

Part 1 (Using internal Dataset)

Given customer web browsing behavior data tracking pages visited, clickstreams and eye-gaze estimates, use machine learning models like regression, random forest and gradient boosting to predict customer conversion probability and expected order value. To guide cross-selling and recommendation engines to maximize customer lifetime value.

The models can ingest features like pages visited, browsing depth, clicked products and gaze time as indicators of purchase intent. Training on historical data tagged with actual sales allows learning complex non-linear patterns that estimate likelihood of closing the sale. A/B experiments can refine model accuracy.

There might also be a relationship between the type of credit card that customers are using and the product they buy. Notably, some credit card companies offer promotions, so there might be specific products that are targeted at certain customers who may be using a promotional offer and are therefore more likely to buy. Given past customer purchases joined with page views and credit card info, use time series clustering to reveal customer segments exhibiting similar purchase patterns over time. The taxonomy of buying behaviors can guide customized recommendations and promotions timed for peak demand cycles.

Classification algorithms such as KNN are useful in identifying similar customers. Clustering algorithms such as K-means may be useful in identifying certain credit card customers who may be purchasing based on an unknown trend. From these algorithms products can be recommended accordingly. Logistic regression can be useful in the estimation of the relationship between time on a web page and the likelihood of making a purchase.

Part 2 (Combining datasets)

A.

Data sets #1 and #2 could be valuable data for a property management company that is doing client screening. So, in this case, "the company" might be able to compile and sell this data to a property management company.

Oftentimes, property managers/landlords may look at applicant's credit scores or rental history, but neither provides a thorough analysis of potential tenants spending habits, wealth or debt management. College or university attended, major or majors, and financial net worth may provide some insight on tenant's income stability. But while they may want to prevent discrimination against any applicant, they should look at these variables from a broader context of their financial net worth or debt management including from data set #2 their credit card habits, how much they owed, how much was paid and whether the person was in default.

Logistic regression can be useful in identifying the relationship between college education, majors, and debt management and the probability that a potential client may stop paying rent. Or Support vector machine and k-means classifiers can be useful in classifying customers who may or may not fail to pay rent.

Alternatively, Given applicant personal attributes like college, major, interests along with historical payment information on rents and credit accounts,
Use logistic regression to estimate probabilistic scores of on-times rental payment linked to educational status, past payment reliability and income proxies.
To construct differentiated tenant reliability profile reports for property management clients to minimize income loss through scientifically rigorous applicant risk scoring.

Where the features are:

- CollegeTier: Elite/High-tier, Medium, Basic
- CreditScore: Numerical score
- MonthlyRentBurden: Percent of income spent on rent
- NumLatePayments: Count in last 2 years

And the target is binary on-time or late rent payment.

The logistic regression model will estimate the probability of on-time payment for new applicants given those attributes.

B.

Another way to monetize these datasets pertains to segmenting. Given alumnus personal attributes and interest data combined with credit bureau payment information, use association rule mining to identify affiliate marketing opportunities. This can uncover what alumni segments aligned to certain majors, regions and industries are most responsive to specific sports and hobby promotions.

To selectively target high response groups for campaigns, maximizing customer acquisitions while minimizing wasted marketing spend. Statistical strength of rules can prioritize the best customer-product fits to monetize data relationships.

Conclusion

In the first part, the focus lies on customer behavior analysis using machine learning models. Leveraging browsing behavior data, models like regression, random forest, and gradient boosting predict customer conversion probabilities and expected order values. These models consider various indicators of purchase intent such as pages visited, click depth, and gaze time. Additionally, integrating credit card data allows exploration of relationships between card types, purchased products, and promotional offers. Segmentation techniques further refine this by identifying customer segments with similar purchase patterns, enabling tailored recommendations and timed promotions. The utilization of classification and clustering algorithms assists in customer identification and predicting purchase likelihood based on web interactions.

Moving to the second part, the analysis delves into the potential of these datasets for property management. Beyond traditional screening metrics, the data offers insights into applicants' education, majors, financial standing, and credit card habits, providing a deeper understanding of potential tenants' reliability in paying

rent. Machine learning models, particularly logistic regression, SVM, and k-means, are suggested to predict rent default probabilities based on educational background and debt management indicators. Moreover, these models help in estimating on-time rental payments using applicant attributes, offering comprehensive tenant reliability profiles for property management clients. Features like College Tier, Credit Score, Monthly Rent Burden, and Late Payments become key in predicting rent payment behaviors.

Finally, the analysis suggests another avenue for monetization through segmentation techniques. By combining alumni attributes, interests, and credit bureau information, association rule mining identifies affiliate marketing opportunities. This approach uncovers responsive alumni segments aligned with majors, regions, and industries, revealing preferences for sports and hobbies. Subsequently, prioritizing these high-response customer segments for targeted marketing campaigns minimizes marketing expenses while maximizing customer acquisition. Overall, the analysis presents a multifaceted approach to extract value from the retailer's datasets, encompassing customer behavior analysis, property management insights, and segmentation strategies for targeted marketing initiatives.