

Analytics Project

4/16/2020

Pick a project for which you think at least three different Analytics models might have been combined to create the solution.

Write a short report describing your answers

I picked a project on increasing sales of offers to existing customers of Cigna insurance company.

The following is a synopsis of the project from the website.

“Can big data analysis increase sales?”

*As a life and health insurer, Cigna focuses on its affinity business, where it collaborates with its partners to serve the customer's need for the right insurance product and services. **Cigna regularly updates the customer database to improve services for existing customers.**”*

*“The CVM department uses SAS to design effective campaigns that **improve response rates, determine the precise profile based on personal data, design a contact strategy with tailor-made offers, and optimize results.**”*

https://www.sas.com/en_us/customers/cigna-indonesia.htmlexisting customers

I will tackle the project in the following sequence:

- 1. How the models would be combined**
- 2. What specific data might be needed to use the models**
- 3. How the data might be collected**
- 4. How often the data might need to be refreshed and the models re-run**
- 5. My conclusion**

Models used to create the solution

1. Support Vector Machine Model
2. Logistic regression Model
3. K Means Clustering Algorithm
4. Optimization Model

1.How models would be combined

Support Vector Machine Model

Use the Support Vector Machine Model to classify the existing Cigna customers as either having:

1. 1 Cigna product
2. 2 or more Cigna products

Logistic Regression

Given: Output from the Support Vector Machine Model on customers that have 1 Cigna product

Use : Logistic regression

To :

1. Set a threshold to predict the probability of a customer buying a second Cigna product or not
2. Classify a customer as yes/no to buying another insurance product from Cigna

K Means Clustering Algorithm

Given: Output from the SVM model and Logistic regression model respectively

Use: K Means Algorithm

To : Sort the customers that have 1 Cigna product and are likely to purchase another product from Cigna into cluster and then target each cluster of customers with targeted messaging or advertising

Optimization

Given :

1. Output from Logistic regression model on the number of customers with one product that will probably buy another product
2. Cluster output from the K Means Algorithm

Use: Optimization

To: balance cost of targeted marketing to customers that have one product with increase in sales i.e. those customers purchasing additional products

n = Number of customers with 1 product predicted to purchase additional Cigna products

p_i = The expected purchase of at least 1 more product by customer

A = Targeted marketing budget

m_i = Minimum amount of targeting marketing cost for each customer

M_i = Maximum amount of targeting marketing cost for each customer

Variables

a_i = money spent target advertising to customer i

b_i = 1 if customer is targeted for marketing, 0 if not

Constraints

$$\sum a_i \leq A$$

$$a_i \geq 0 \text{ for all customers } i$$

$$a_i \leq A b_i \text{ for all customers } i$$

$$a_i \geq m_i b_i \text{ for all customers } i$$

$$a_i \geq M_i b_i \text{ for all customers } i$$

Objective Function

Maximize difference between expected purchase of at least 1 more product by customers and money spent on targeted marketing to customers

$$\text{Maximize } \sum_{i=1}^n p_i - \sum_{i=1}^n a_i$$

2.What specific data might be needed to use the models

Data to be used in the Support Vector Machine Model

1. Customer credit score
2. Customer age
3. Customer income
4. Customer education level
5. Customer medical conditions
6. Customer history of insurance claims
7. Customer payment history
8. Customer satisfaction with service
9. Customer satisfaction with product
10. Length of business relationship with Cigna
11. Number of products' customer has with Cigna (1 or 2 or more)

Data to be used in Logistic Regression

Output from SVM model of customers with 1 Cigna product.

Use the following data from that specific group of customers:

1. Length of business relationship with Cigna
2. Customer credit score
3. Customer age
4. Customer income
5. Customer medical conditions
6. Customer history of insurance claims
7. Customer payment history
8. Customer satisfaction with service
9. Customer satisfaction with product
10. Customer perception of Cigna

Data to be used in the K Means Algorithm

Output from the logistic regression model of customers likely to buy another Cigna product.

Use the following data from that specific group of customers

1. Customer credit score
2. Customer education level
3. Customer age
4. Customer income
5. Customer medical conditions
6. Customer history of insurance claims
7. Type of product customer already has

Data to be used in Optimization Model

1. Output from Logistic regression model on the number of customers with one product that will probably buy another product
2. Cluster output from the K Means Algorithm
3. Targeted marketing budget

3.How the data might be collected

1. Customer demographics from applications for Cigna health and life insurance products.
2. Customer demographics, customers financials, products purchased and sales data from Cigna's database and Cigna partners data bases.
3. Customer health and life insurance claim information from Cigna and their partner companies.
4. Questionnaires and surveys done online or in person, with existing customers to get customer service data, product satisfaction data and company perception.
5. Polls and emails to gather information on customer preferences and opinions.
6. Focus groups to determine interest in a certain product and group perception of the company.

4.How often data might need to be refreshed and the models re-run

The data should be refreshed every 3 months and the models re-run at the same time.

Health and insurance products are a great source of monetary benefit because most people need to have products like health insurance ,long term care insurance and life insurance.

There are periods of time when people are more likely to purchase certain kinds on insurance for example most people begin looking at long term insurance as they get older and start thinking of how they will be able to take care of themselves as they age.

Life insurance purchases are more common in people who are married and have children versus single and childless people in their early twenties.

There are also times of the year when people are more likely to make insurance purchasing decisions. For example, a lot of people make new year resolutions either at the end of the year or beginning of the next year. This would be a good group to capture and extend an offer of an insurance product or an additional insurance product.

Refreshing the data every 3 months is often enough to capture the myriad of people with changing insurance needs but not too often that it would be cost ineffective or not capture a significant change in the data.

5.Conclusion

In conclusion my project addressed the following statement from the synopsis:

*“The CVM department uses SAS to design effective campaigns that **improve response rates, determine the precise profile based on personal data, design a contact strategy with tailor-made offers, and optimize results.**”*

Improve response rates by using SVM to find the customers who have 1 product and then using logistic regression to predict those customers in that group that would be more likely to buy another product. This would reduce time and resources wasted marketing to the customers who would likely not purchase another product, therefore improving response rates to the sales marketing .

Determine the precise profile based on personal data by using the K Means algorithm to cluster the customers likely to buy another insurance product into different clusters based on different attributes like income, education level, product type customer already has, age, credit score and history of previous claims.

Design a contact strategy with tailor-made offers using the cluster information from the K Means algorithm to target advertising to each of the clusters based on buying motivation .One cluster might purchase an additional product based on past medical history and future need for example long term care insurance whereas another cluster might purchase a new product based on asset protecting or leaving a legacy ,for example life insurance. Knowledge of the clusters and their potential buying motivations would enable Cigna to make tailor-made offers to each cluster.

Optimize results using an optimization model to maximize the difference between resources(time and money) spent on targeted marketing versus the increase in sales from customers purchasing additional products.

