

Introduction

In insurance marketing, a well-designed personalized interaction with customers is extremely important. The agents who are fighting for new clients and trying to retain existing clients like to develop a personal relationship with their clients which would ultimately have a positive effect on their baseline. Consumers' priorities for different insurance products change over time as they move through different stages of their lifecycles. For an insurance company it is therefore important to understand how particular products in its portfolio appeal to different customers as they progress through their lifecycles so appropriate up-sell and cross-sell strategies can be initiated. Based upon insurance usage data and socio-demographic data derived from zip codes, JMP® Pro 10 is used to Profile the existing buyers and also predicted the future buyers of the mobile home insurance product.

Methods

Data Preparation

The survey based dataset consists of 9000 observations and 86 variables. It is a multivariate dataset having no missing values. The target variable is a binary variable explaining whether a customer bought an insurance product.

Following steps were taken during data preparation and exploration:

- Identification of multi-collinear data using multivariate methods (Ref. fig.1)
- Outlier analysis using Mahalanobis distance statistic
- Double-standardization of data using JSL for effective clustering

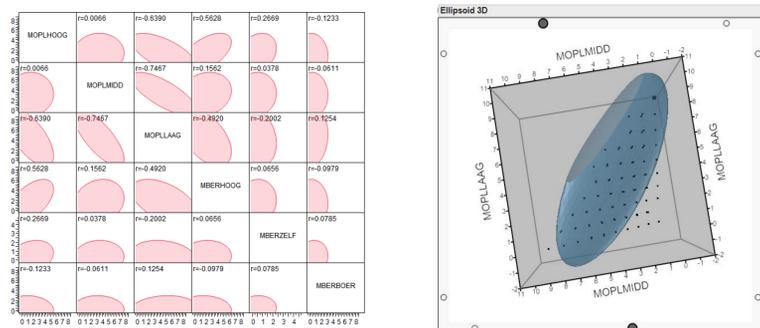


Fig. 1: Scatterplot matrix and Ellipsoid 3D Plot showing correlation amongst variables

Clustering, Segment Profiling and Predictive Modeling

To segment customers of the insurance company in distinct groups several clustering methods were applied on the double-standardized data. The newly formed segments were then profiled and characteristics of each segment were understood. After segments were created, different predictive models such as logistic regression and neural networks were applied and compared to predict potential buyers of mobile home insurance product.

Following steps were taken during this phase:

- Creation and evaluation of clusters using hierarchical and k-means clustering methods (Ref, Fig. 2)
- ANOVA testing to compare means of variables across segments (Ref. Fig. 3)
- Building predictive models using Logistic Regression and Neural networks (Ref. Fig. 4)

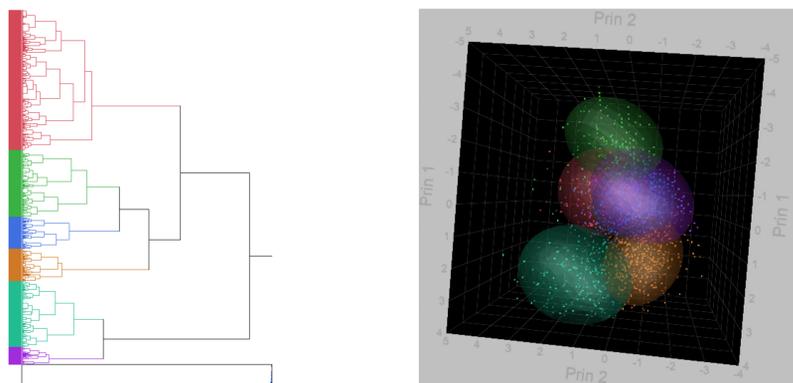


Fig. 2: Clusters created using hierarchical and k-means clustering methods

Results

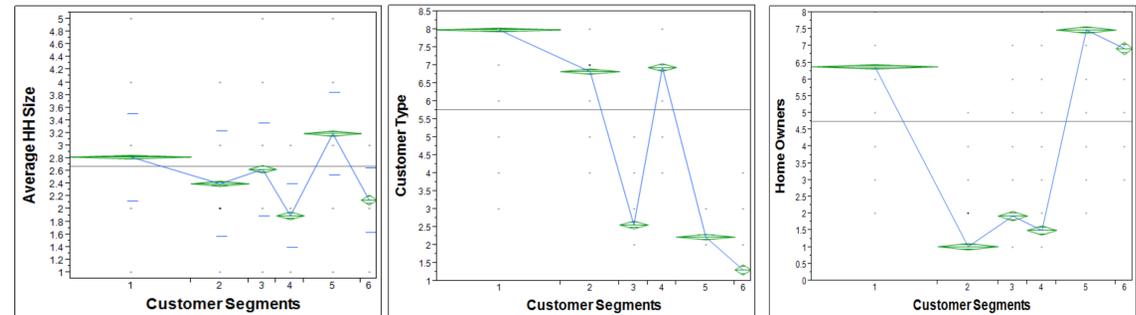


Fig. 3: Profiled Segments using hierarchical clusters

Segment	Description
1	Group of people in their early 50s having families and grown up children. Living in their own house.
2	Senior citizens who are wealthy, having a high discretionary income and living in a rented house.
3	Young people in their early 30s, living well in a rented house. This segment mostly comprised of unskilled labor who is not highly educated.
4	Well educated people from an average family having medium to high income. They also have high purchasing power.
5	Mostly farmers having grown up children. Not very educated having low to medium income.
6	Young couples or singles having high purchasing power and maintaining a high status in society. They are middle level managers having their own house.

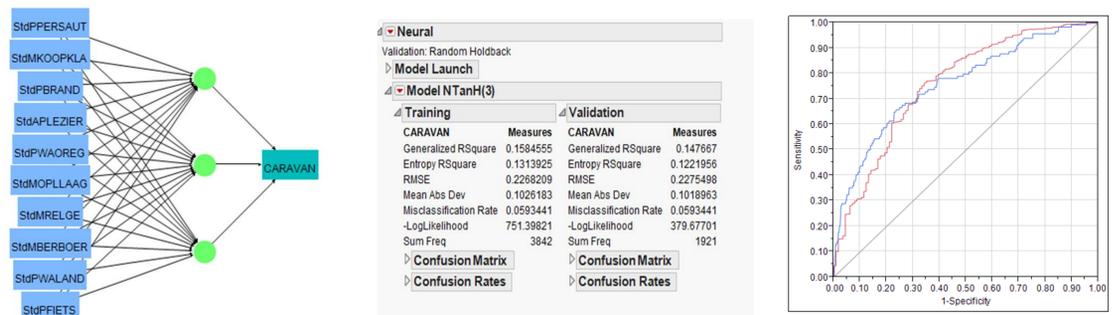


Fig. 4: Neural network performance on the standardized data

Conclusion and Discussion

- Like discussed before in previous SAS GF paper by Pagolu et al. (2011), double standardization of data improves performance of clustering algorithms. It has been identified that double-standardization also improves the accuracy of predictive models
- Neural network model was performing better than logistic regression, hence it was selected as a final model
- Factors such as age, marital status, purchasing power and type of previous insurance policies are important in deciding whether a person would buy mobile home insurance product
- From our analysis it was also observed that customers aged between 35-45 years, who are married and hail from higher income group with better purchasing power are more likely to buy the insurance product

Reference

- Pagolu, & Chakraborty, (2011). Eliminating Response Style Segments in Survey Data via Double Standardization Before Clustering. SAS® Global Forum
- Sarkar & Nakkeeran, (2012). Data Standardization for Enhanced Outcomes in Cluster Analysis Using JMP®

Acknowledgements

- P. van der Putten and M. van Someren (eds) . CoLL Challenge 2000: The Insurance Company Case. Published by Sentient Machine Research, Amsterdam
- Prof. Dr. Goutam Chakraborty, founder of SAS and OSU Business Analytics Program at Oklahoma State University, for his continued support and guidance