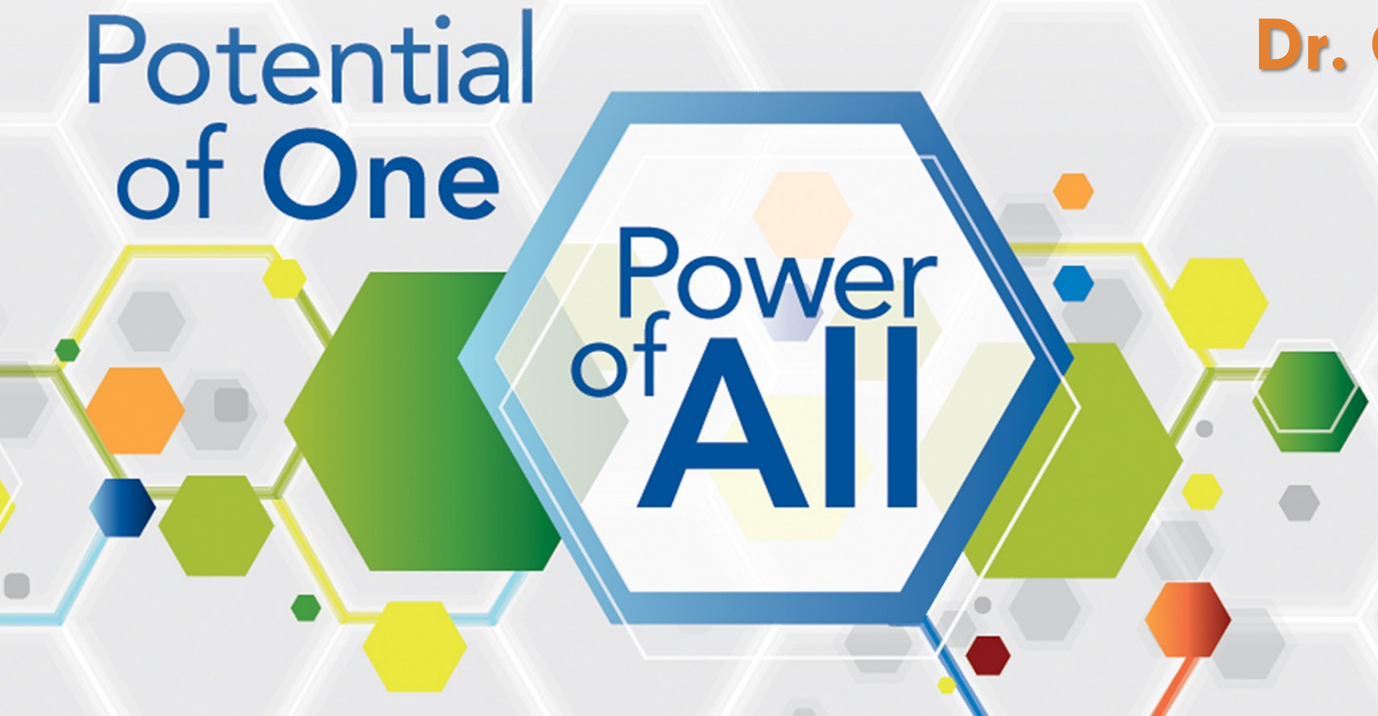
Predicting a potential child smoker using SAS® Enterprise Miner 12.1

Jin Ho Jung
Gaurav Pathak
Dr. Goutam Chakraborty



Predicting a potential child smoker using SAS® Enterprise Miner 12.1

Jin Ho Jung[1], Gaurav Pathak[2], Dr. Goutam Chakraborty[1]
[1]Department of Marketing, Oklahoma State University, Stillwater, OK 74078
[2]Management Information Systems, Oklahoma State University, Stillwater, OK 74078

Introduction

Over the years, there has been a growing concern about consumption of tobacco amongst youth. But no concrete studies have been done to find out what exactly leads children to start consuming tobacco.

This study is an attempt to figure out the potential reasons for the same. Through our analysis we have also tried to build a model to predict whether a child would smoke next year or not .

This study was based on the National Youth Tobacco Survey. It provides national data key to design, implement, and evaluate comprehensive tobacco prevention and control programs. Base SAS® 9.3 and SAS® EM 12.1 are used to predict the best model and find significant factors that affect tobacco usage among young children.

Such findings suggest how public policy makers can effectively campaign for tobacco-free youth

Method

- **Data Preparation**
  - Data preparation is of primary importance when it comes to solving a business problem. The survey-based data initially consisted of 18,867 observations and 197 variables.
  - The target variable is a binary variable which indicates whether a child would smoke next year or not.
  - From a pool of 192 potential input variables, 13 important variables were selected (Fig. 1) using variable selection methods, partial least squares and decision tree models.

Variable	Description
Qn3	Grade in which the child is studying.
Qn20	Did anyone ever refuse to sell you cigarettes because of your age?
Qn52	During the past 12 months, did any doctor, dentist, or nurse ask you if you use tobacco of any kind?
Qn68	How many of your closest friends smoke cigarettes?
Qn73	Do you think smoking cigarettes makes young people look cool or fit in?
access_to_tobacco	How did you get your cigarette?
ads_actors	When you watch TV or go to movies, how often do you see actors using tobacco?
ads_fav_brand	What is the name of the cigarette brand of your favorite ad?
different_tobacco	Which of the following tobacco products have you ever tried?

Fig. 1: List of Selected Variables

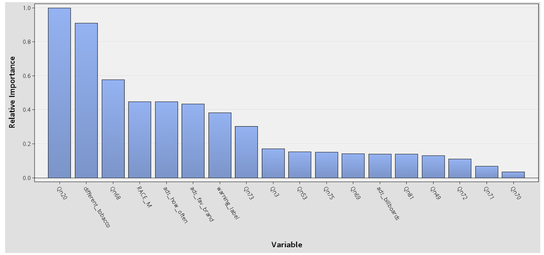


Fig. 2: Chi-square values

- **Predictive Modeling**
  - Once the data was prepared and ready for analysis, it was partitioned into Training, Validation and Testing with a ratio of 50:30:20.
  - As a part of predictive modeling, Decision Trees, Stepwise Logistic Regression Model and Stepwise Logistic Polynomial Regression Model were built based on the previously selected variables. Using Model Comparison Node in SAS®® Enterprise Miner 12.1, competing models were diagnosed and compared with each other.
  - Stepwise Logistic Regression Model outperformed other models (See Fig. 3). With misclassification rate as our selection criteria, stepwise regression model produced Validation misclassification of 0.028497.
  - Factors such as *company of friends, cigarette brand ads, accessibility to the tobacco products,* and passive smoking turned out to be most important predictors in determining a child smoker.

Selected Model	Model Node	Model Description	Target Variable	Valid: Misclassification Rate
Y	Reg	Regression	Target	0.028497
	Tree	Decision Tree	Target	0.0295
	Reg2	Polyomial Regression	Target	0.032109

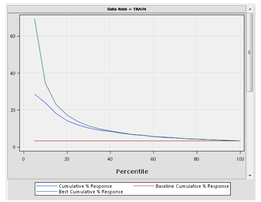


Fig. 3: Stepwise Logistic Regression Model Performance

Discussion

- The purpose of this study is to investigate whether a youth would smoke next year and which factors influence youth to start smoking.
- Tobacco companies' advertising and the number of smoker friends a youth has mostly influence a youth's willingness to smoke in the next year.
- This study gives directions to public policymakers and school teachers to develop and implement tobacco prevention programs. This will help to reduce youths' tobacco use and ultimately prevent illness and death caused by tobacco use.
- We believe that our findings provide insightful information to both public policymakers and school teachers in order to effectively control and campaign youths' tobacco use.

References

- Feighery, E. C., Ribisl, K. M., Schleicher, N., Lee, R. E., & Halvorson, S. (2001).
  Cigarette advertising and promotional strategies in retail outlets.
- U.S. Department of Health & Human Services (2013), Youth & Tobacco
- Wayne W. LaMorte, MD, PhD, MPH, Boston University School of Public Health (2013).

SAS® GLOBAL FORUM

Washington, D.C.

March 23–26, 2014