

Using SAS® Enterprise Miner to Examine the Severity of Nicotine Dependency in India

AUTHORS

Akansha Singh
Postdoctorate,
Centre de recherché en demographie
Universite Catholique De Louvain
Belgium

Himanshu Katyan
Senior Statistician
Inventiv Health
Belgium

Tobacco use in India

- ❑ India has one of the highest tobacco users in the world both in number and relative share.
- ❑ India is one of the fewer countries in the world where prevalence of smoking and smokeless tobacco use are high.
- ❑ Dual use of tobacco (use of both smoking and smokeless tobacco products) also contributes to a noticeable proportion
- ❑ Tobacco use in India is responsible for a large number of premature deaths in India.
- ❑ The majority of smoking related deaths in India occur in the prime working age group of 15–59 years.
- ❑ Smokeless tobacco use is also associated with the increasing risk of cancer.
- ❑ Smokeless tobacco is also highly addictive and causes cancer of the head, neck, oesophagus, besides many oral diseases.

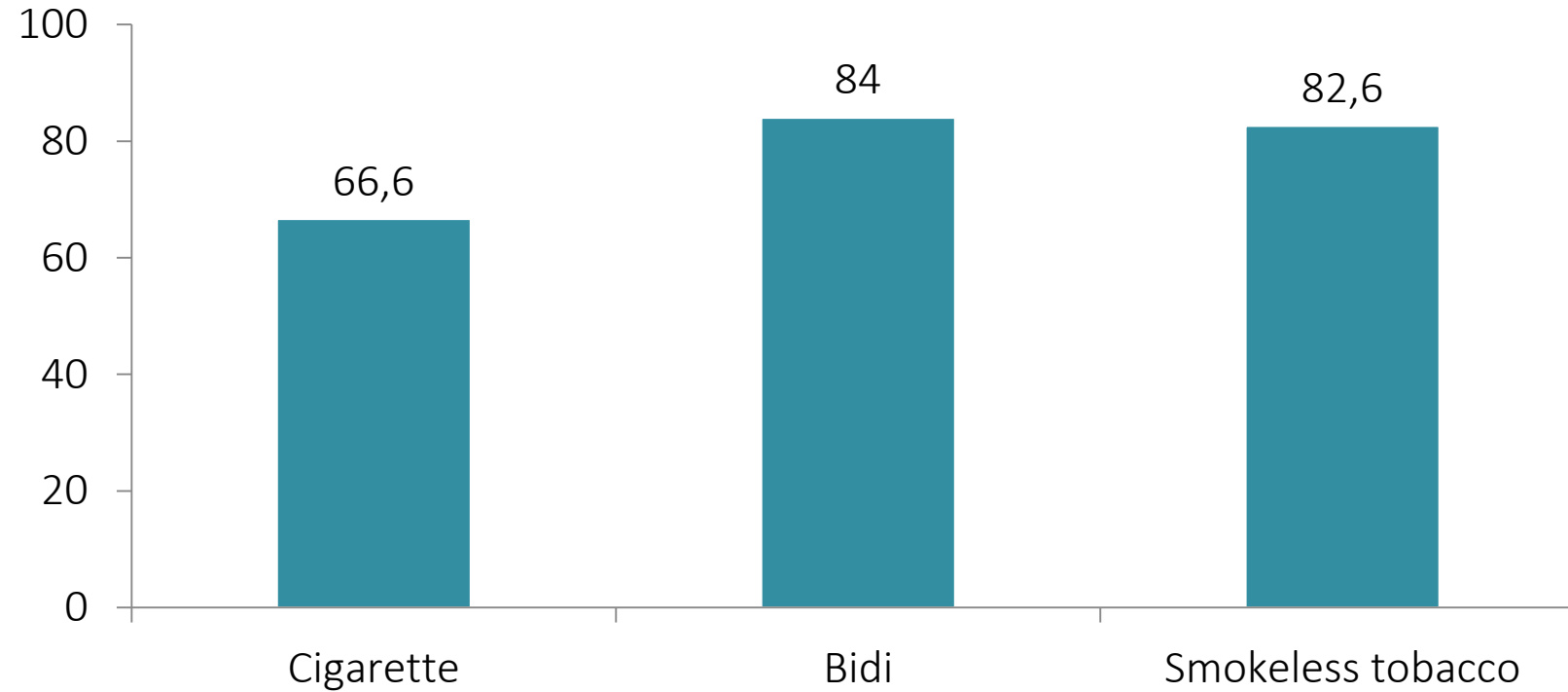
Global Adult Tobacco Survey India (GATS India)

- ❑ GATS India was the first nationwide survey in which electronic handheld devices were used for data collection and management.
- ❑ GATS India is the global standard for systematically monitoring adult tobacco use (smoking and smokeless).
- ❑ 15 Countries Participating in GATS :- Bangladesh, China, India, Indonesia, Pakistan, Philippines, Thailand, Turkey, Vietnam, Brazil, Mexico, Egypt, Poland, Russian Federation, Ukraine.
- ❑ GATS International agencies partnered with The Ministry of Health and Family Welfare, (MoHFW), Government of India to conduct this survey in India.
- ❑ Technical assistance was provided by the Centers for Disease Control and Prevention (CDC), the World Health Organization (WHO), the Johns Hopkins Bloomberg School of Public Health, and Research Triangle Institute International (RTI International).

Tobacco use in India - Findings from GATS India Survey

- ❑ GATS India was conducted in 2009–2010 as a household survey of persons age 15 and above.
- ❑ A total of 69,296 interviews were completed among which 33,767 and 35,529 were of males and females respectively.
- ❑ Out of all completed interviews, 41,825 interviews were conducted in rural areas and 27,471 interviews in urban areas.
- ❑ One-third (35%) of adults in India use tobacco in some form or the other.
- ❑ Among them 21% adults use only smokeless tobacco, 9% only smoke and 5% smoke as well as use smokeless tobacco.
- ❑ More than 75 % of tobacco users, both smokers as well as users of smokeless tobacco are daily users of tobacco.

PERCENT OF SMOKERS/USERS OF SMOKELESS TOBACCO WHO REPORT TO USE IT ON A DAILY BASIS



Objectives

- ❑ To examine the extent of nicotine dependency among adults in India.
- ❑ To examine the factors effecting nicotine dependency among adults in India.

Data and Methodology

Data

- Global Adult Tobacco Survey 2009-2010

Outcome Indicators

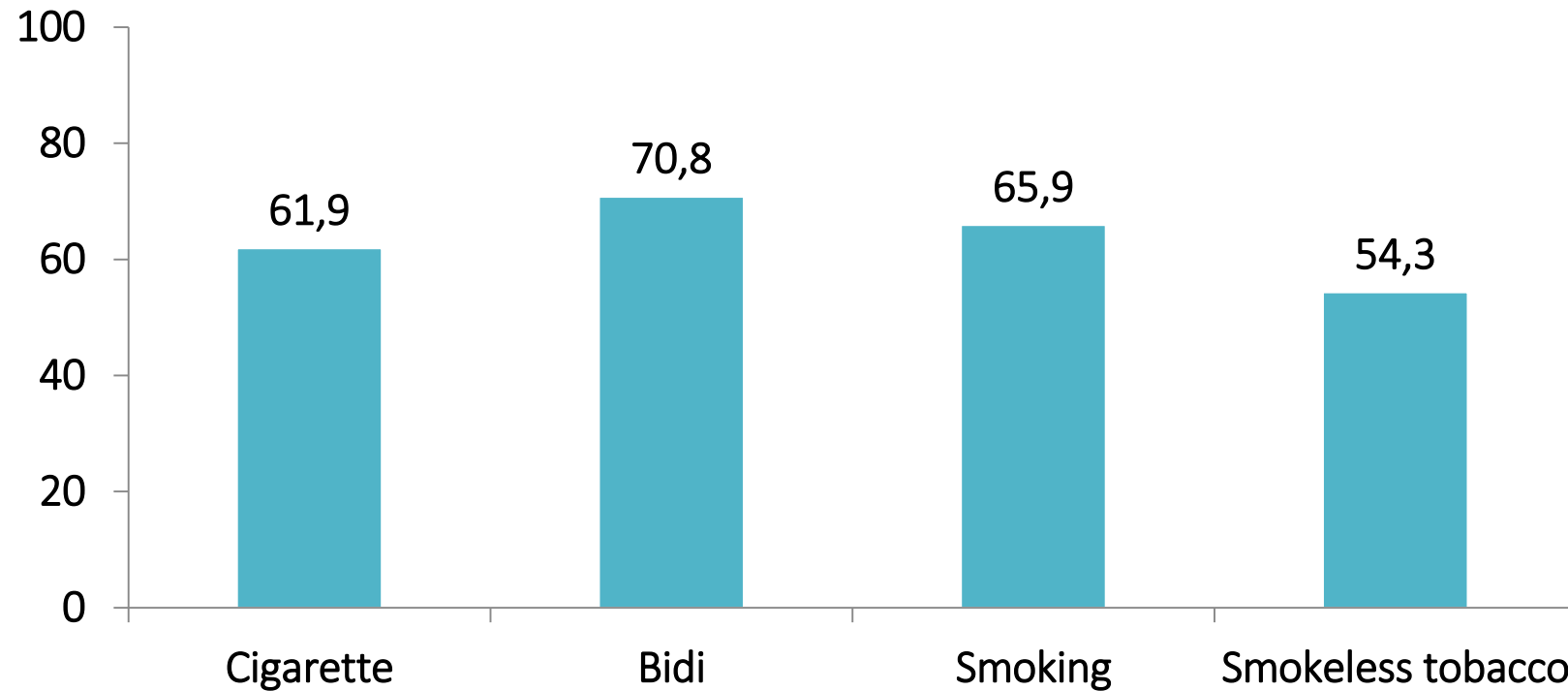
- Daily smoker whose first use of smoking products is within 30 mins of waking up (0 1)
- Daily smokeless tobacco user whose first use of tobacco is within 30 mins of waking up (0 1)

Statistical Methods

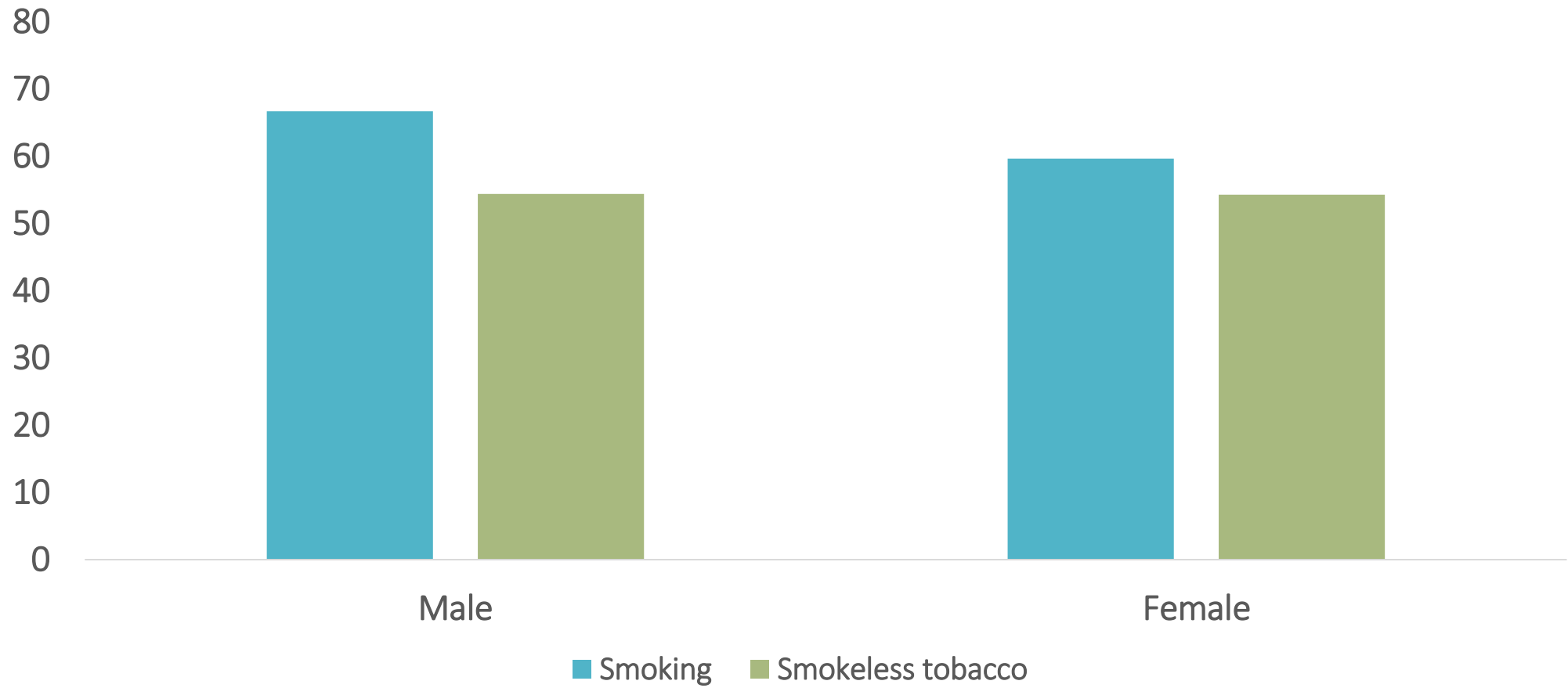
- Bi-Variate Analysis from BASE SAS programming
- SAS[®] Enterprise Miner - Decision Tree analysis

RESULTS

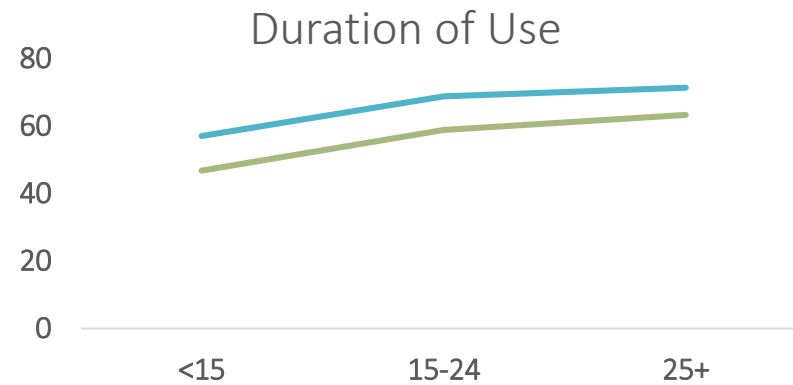
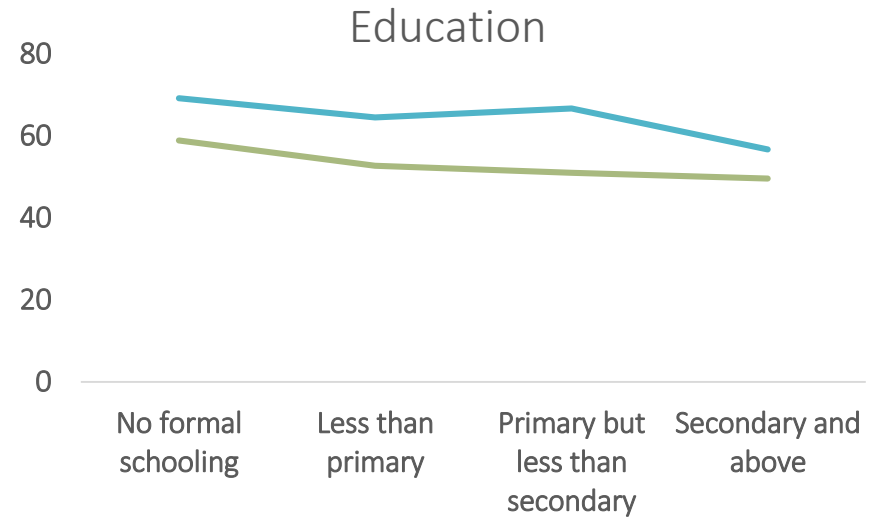
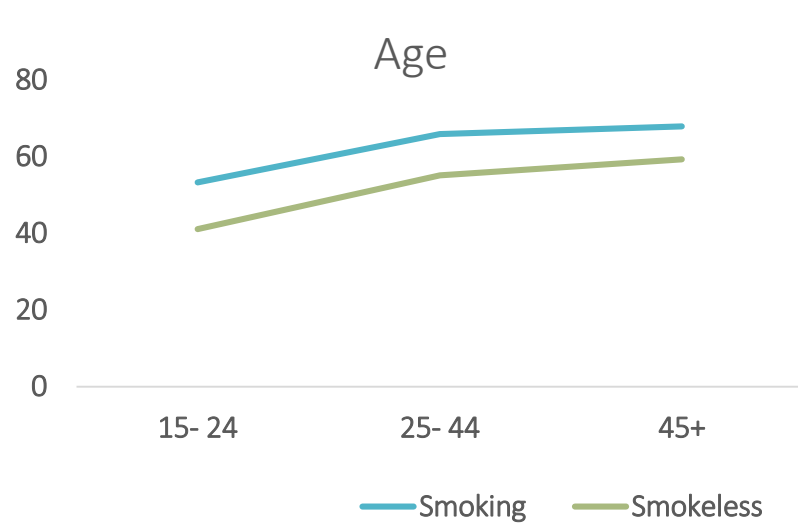
PERCENT OF DAILY TOBACCO USERS WHO TAKE FIRST TOBACCO OF THE DAY WITHIN 30 MINUTES AFTER WAKING UP



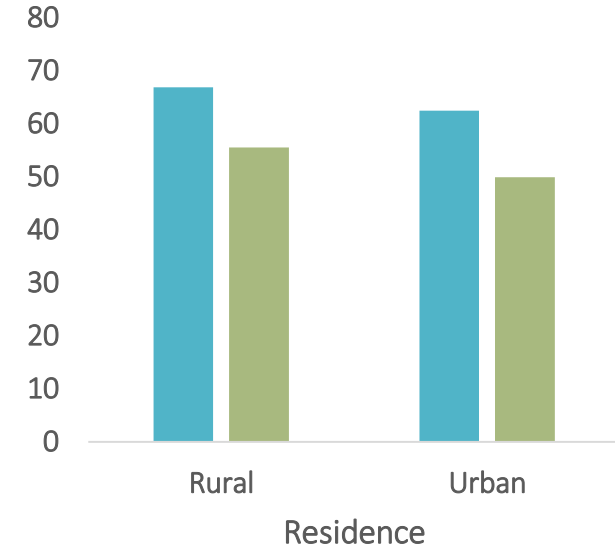
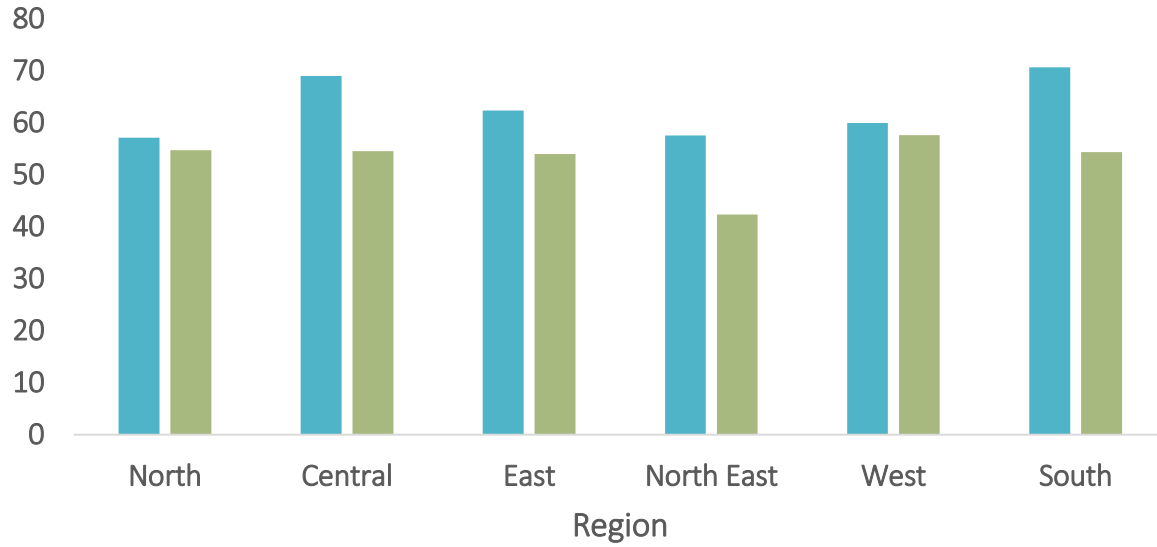
Nicotine Dependency For Males And Females



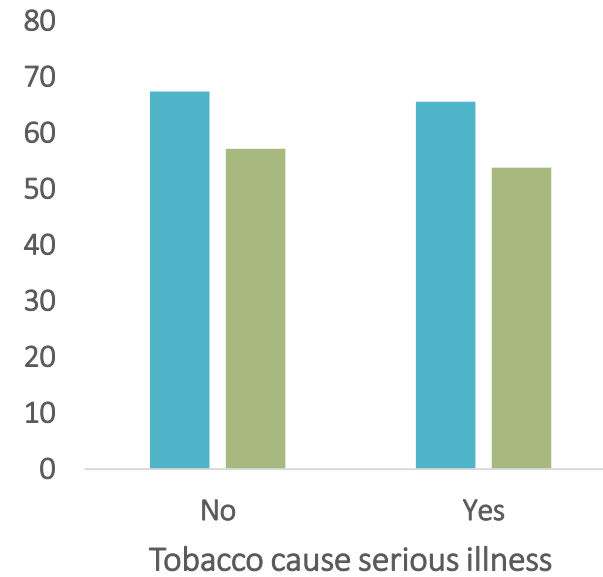
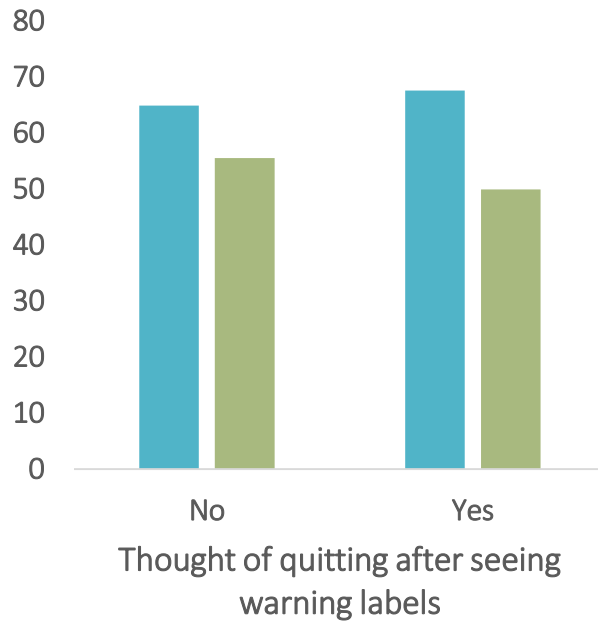
Nicotine Dependency by Age, Education And Duration Of Use



VARIATION IN NICOTINE DEPENDENCY BY PLACE, QUITTING THOUGHT AND HEALTH AWARENESS



Smoking Smokeless



SAS® ENTERPRISE MINER, Decision Tree analysis

Why Decision Tree

- Decision trees involve splitting the data into groups by successively dividing the data into subgroups based on empirically derived associations between the response (target) and one or more predictor variables.
- Decision trees attempt to find this association in a group of observations that form a data set.
- When a set of values is identified as having a strong relationship to a target value, all of these values are grouped in a bin that becomes a branch on the decision tree.
- Decision trees are powerful and popular tools for classification and prediction.
- Non-parametric and non-linear method
- Perform classification without much computation
- Can handle continuous and categorical variables
- Provides a clear indication of which fields are most important for prediction or classification

Analysis Approach

- ❑ Variable selection node is used to find out the most important input variables to predict the target variable.
- ❑ The data is partitioned into training (60%) and validation (40%) data set.
- ❑ Decision models are built using decision node properties
- ❑ Leaf size- Based on Average Square error
- ❑ Maximum Depth - 6
- ❑ Maximum number of Branch – 5
- ❑ Significance level for the model – 0.05

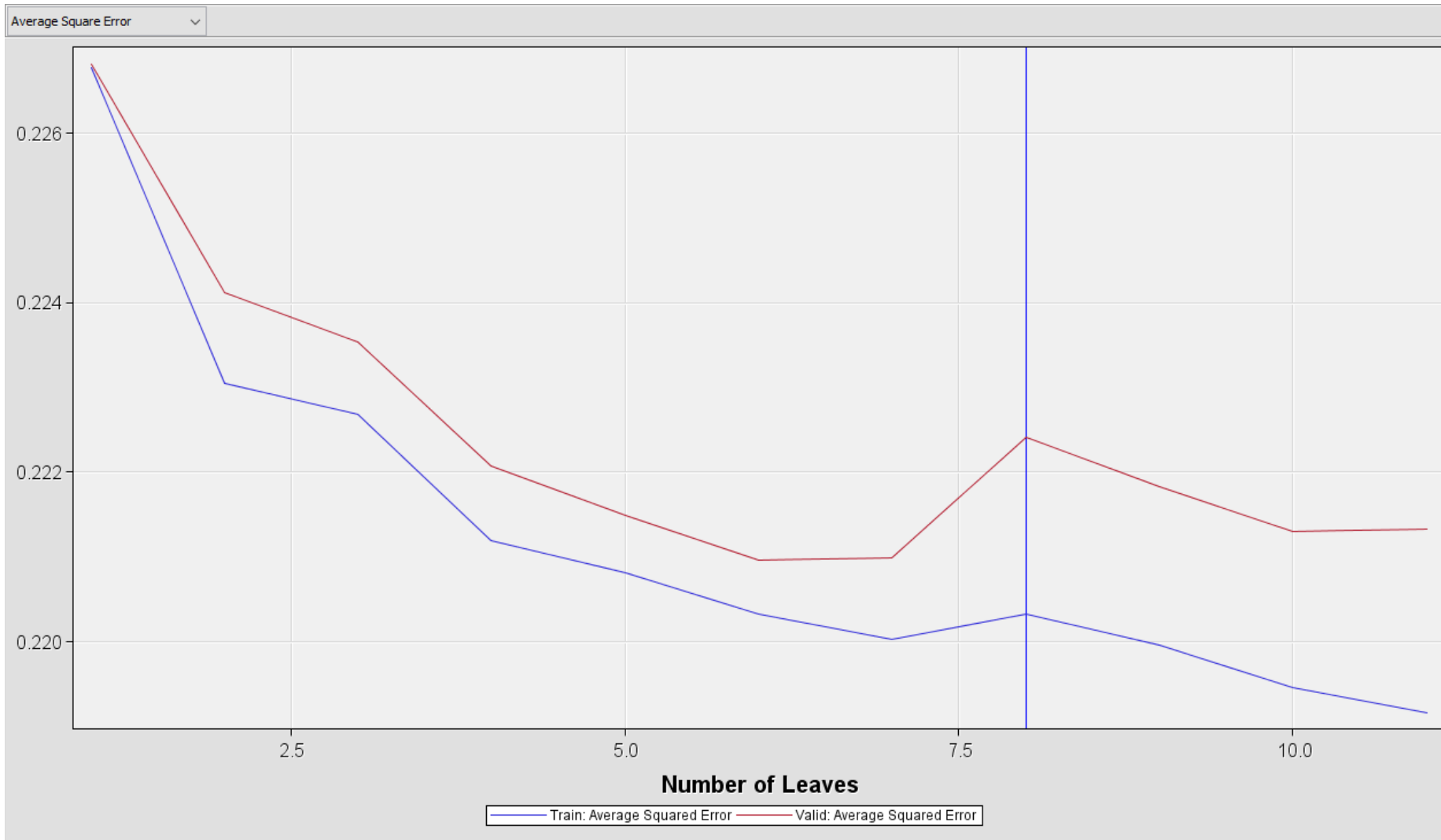
Selected variables list for decision tree model

Smoking	Smokeless
Age	Age
Sex	Sex
Education	Education
Occupation	Occupation
Duration of smoking	Duration of smokeless use
Region of India	Region of India
Residence	Residence
Thought of quitting after seeing warning on smoking products	Thought of quitting after seeing warning on smokeless products
Smoking cause serious illness	Smokeless cause serious illness
Number of members in household	Number of members in household
Smoking is allowed in home	

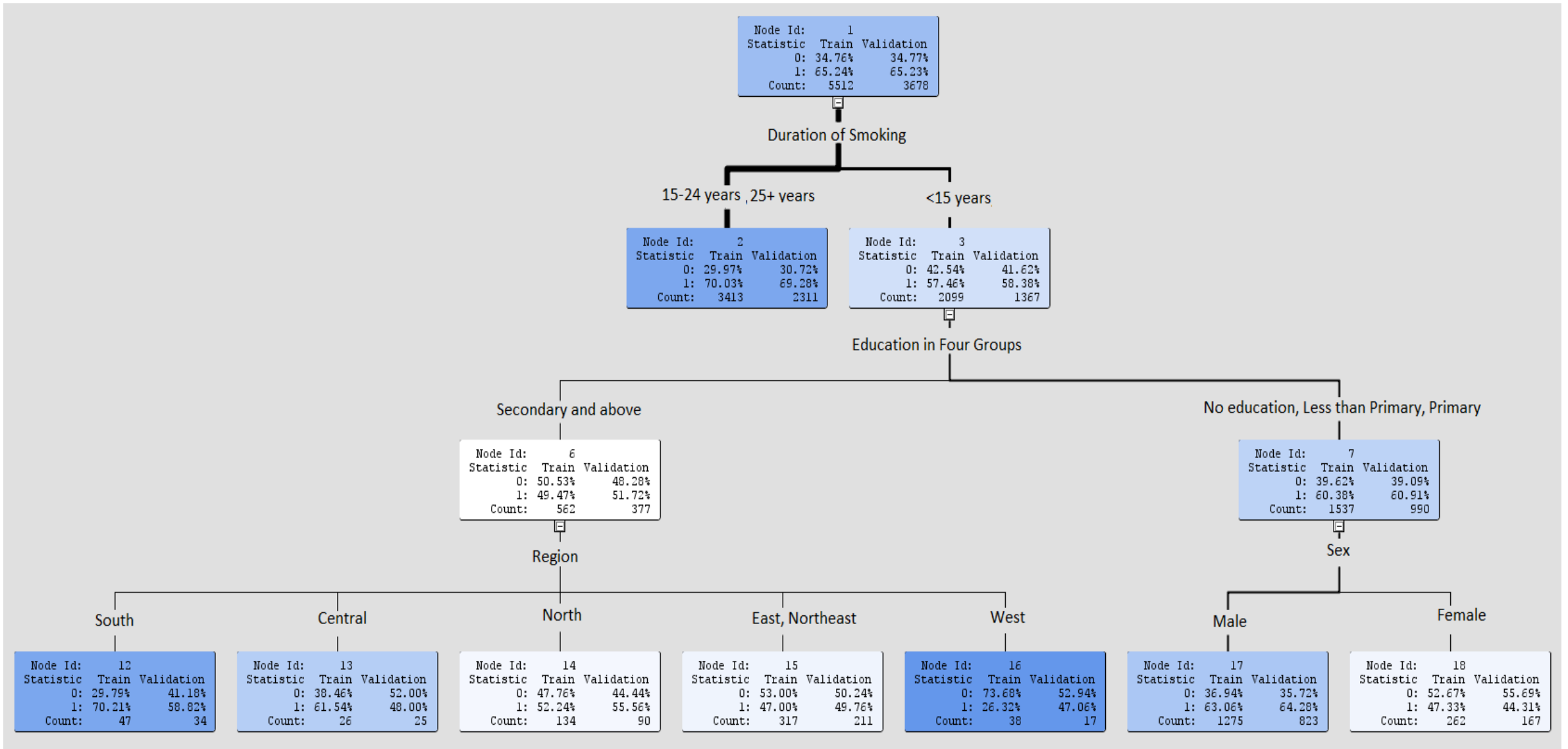
Fit Statistics of Decision Tree Models

	Smoking		Smokeless tobacco	
	Train	Validation	Train	Validation
Sum of Frequencies	5512	3678	8013	5344
Misclassification Rate	0.34	0.34	0.43	0.42
Minimum Absolute Error	0.74	0.74	0.71	0.71
Sum of Squared Error	2428.88	1636.06	3875.09	2593.71
Average Squared Error	0.22	0.22	0.24	0.24
Root Mean Squared Error	0.47	0.47	0.49	0.49

Decision Tree Leaves Analysis - Smoking



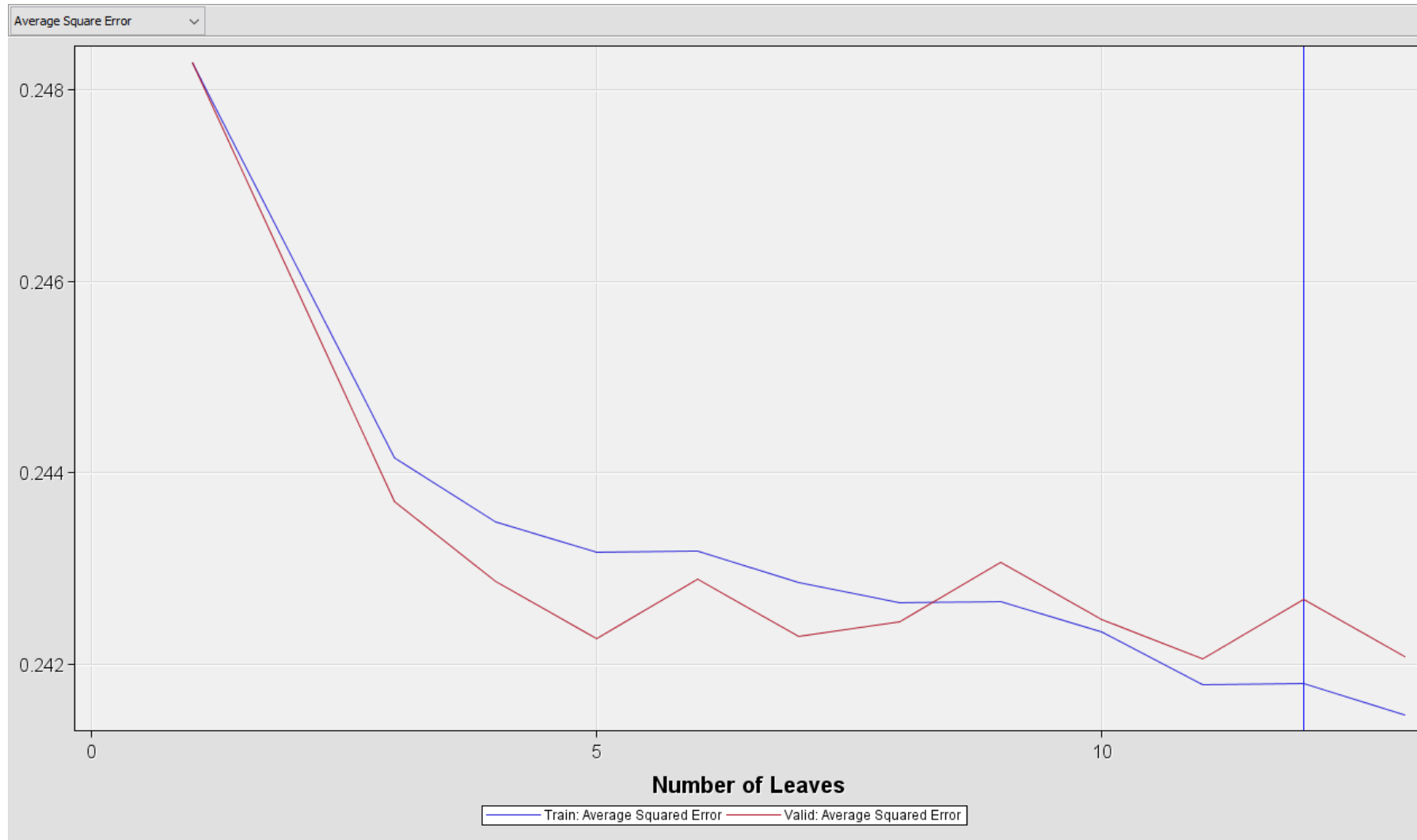
Decision tree - Smoking Dependency



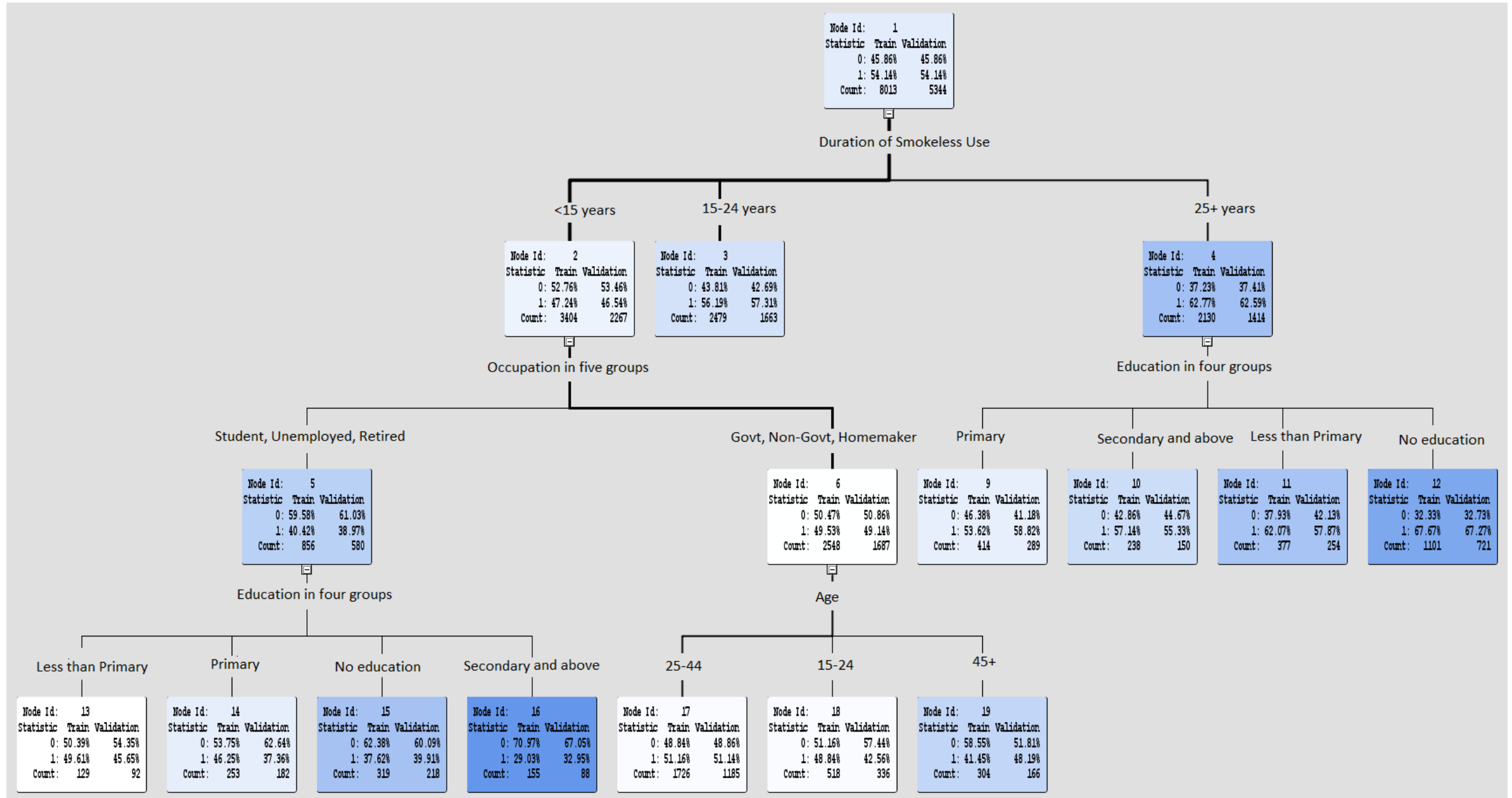
Main results – Decision Tree Smoking

- ❑ Duration of smoking first characterizes the smoking dependency (0-1) response.
- ❑ 70% of the long duration smokers (more than 15 years) are more addicted to towards tobacco consumption compare to the short duration smokers (less than 15 years).
- ❑ Under the short duration smokers node in the decision tree, short duration smokers could be further categorized based on educational attainment.
- ❑ Low educated are more addicted (60.38%) compare to higher educated (49.47%) within short duration smokers.
- ❑ Low educated smokers are further classified into significant male-female subgroup and it shows higher dependency among males smokers (63%).
- ❑ National region also further sub-classified highly educated smokers.
- ❑ High educated short duration smokers mainly from the South and Central region have higher dependency towards smoking.

Decision Tree Leaves Analysis – Smokeless Tobacco



Decision Tree – Smokeless Tobacco Dependency



Main results – Decision Tree Smokeless Tobacco

- ❑ Duration of consuming tobacco in smokeless form play most important role in classifying smokeless tobacco users.
- ❑ Education attainment play prominent role in classifying smokeless tobacco users for more than 25 years.
- ❑ 67.6 % of the long duration illiterate smokers are addicted towards tobacco consumption as compared to the 57% of the higher educated users.
- ❑ Short duration users were also further classified into significant further subgroups by occupational categories, age group and education.
- ❑ Government and non-government employee, self-employed are more addicted than students, retired or unemployed people.
- ❑ Middle aged adults (25-44 years) working in government/non-government sector or self employed have higher dependency than older or young adults.

CONCLUSIONS

- ❑ By using decision tree model, it is possible to identify the combinations of factors that constitute the highest (or lowest) risk for a condition of interest - nicotine dependency in our case and its relative importance.
- ❑ Duration of tobacco use is the most important risk factor of nicotine dependency in India i.e. higher duration of tobacco use leads to higher dependency of smoking or smokeless tobacco.
- ❑ Low educated male smokers are the another high risk category of smokers with significantly higher smoking dependency.
- ❑ Illiterate long duration (25+ years) smokeless users have the highest dependency for smokeless tobacco.
- ❑ Short duration middle aged (25-44 years) users working in government and non government employees have significantly higher dependency towards smokeless tobacco.

Thank You