

SAS® GLOBALFORUM 2015

The Journey Is Yours

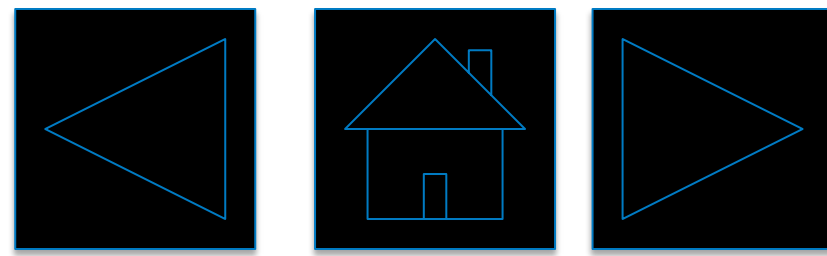
Discovering Personality Type
through Text Mining in SAS
Enterprise Miner 12.1

Deloitte & Touche LLP



Discovering Personality Type through Text Mining in SAS Enterprise Miner 12.1

Deloitte & Touche LLP



Abstract

Data scientists and analytic practitioners have become obsessed with quantifying the unknown. Through text mining third-person posthumous narratives in SAS Enterprise Miner 12.1, we can measure tangible aspects of personalities based on the broadly accepted big-five characteristics: extraversion, agreeableness, conscientiousness, neuroticism, and openness. These measurable attributes are linked to common descriptive terms used throughout our data to establish statistical relationships. The data set contains more than 1,000 obituaries from newspapers throughout the United States with individuals who vary in age, gender, demographic, and socio-economic circumstances. In our study, we leveraged existing literature to build the ontology utilized in the analysis; this literature suggests that a third person's perspective gives insight into one's personality, solidifying the use of obituaries as a source for analysis. We statistically linked target topics such as career, education, religion, art, and family to the five characteristics. With these taxonomies, we developed multivariate models to assign scores to predict an individual's personality type. With a trained model, this study has implications to predict an individual's personality, allowing for better decisions on human capital deployment. Even outside the traditional application of personality assessment for organizational behavior, the methods used to extract intangible characteristics from text allows us to identify valuable information across multiple industries and disciplines.

Methodology

DEFINE

- Leverage academic sources to derive a taxonomy for different personality types
- Identify broad topics to build the sampling frames for various social demographics

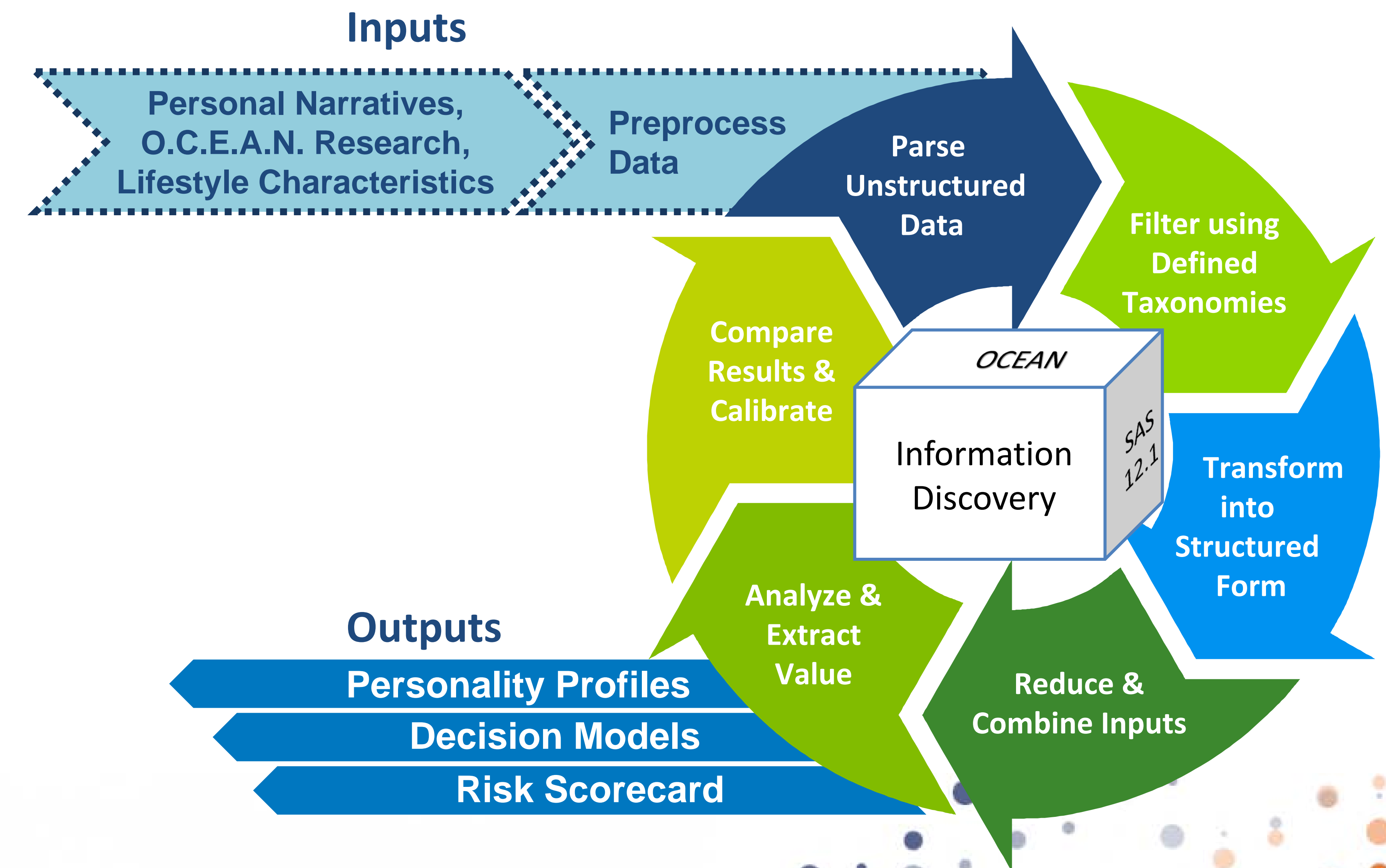
MEASURE

- Quantify intangible characteristics by building target based term weights
- Assess relative importance of intangible characteristics for each particular individual

ANALYZE

- Evaluate statistical relationships between personality traits and social demographic classifications
- Employ multivariate models to predict the magnitude of a personality trait in a given individual

Overview



Discovering Personality Type through Text Mining in SAS Enterprise Miner 12.1

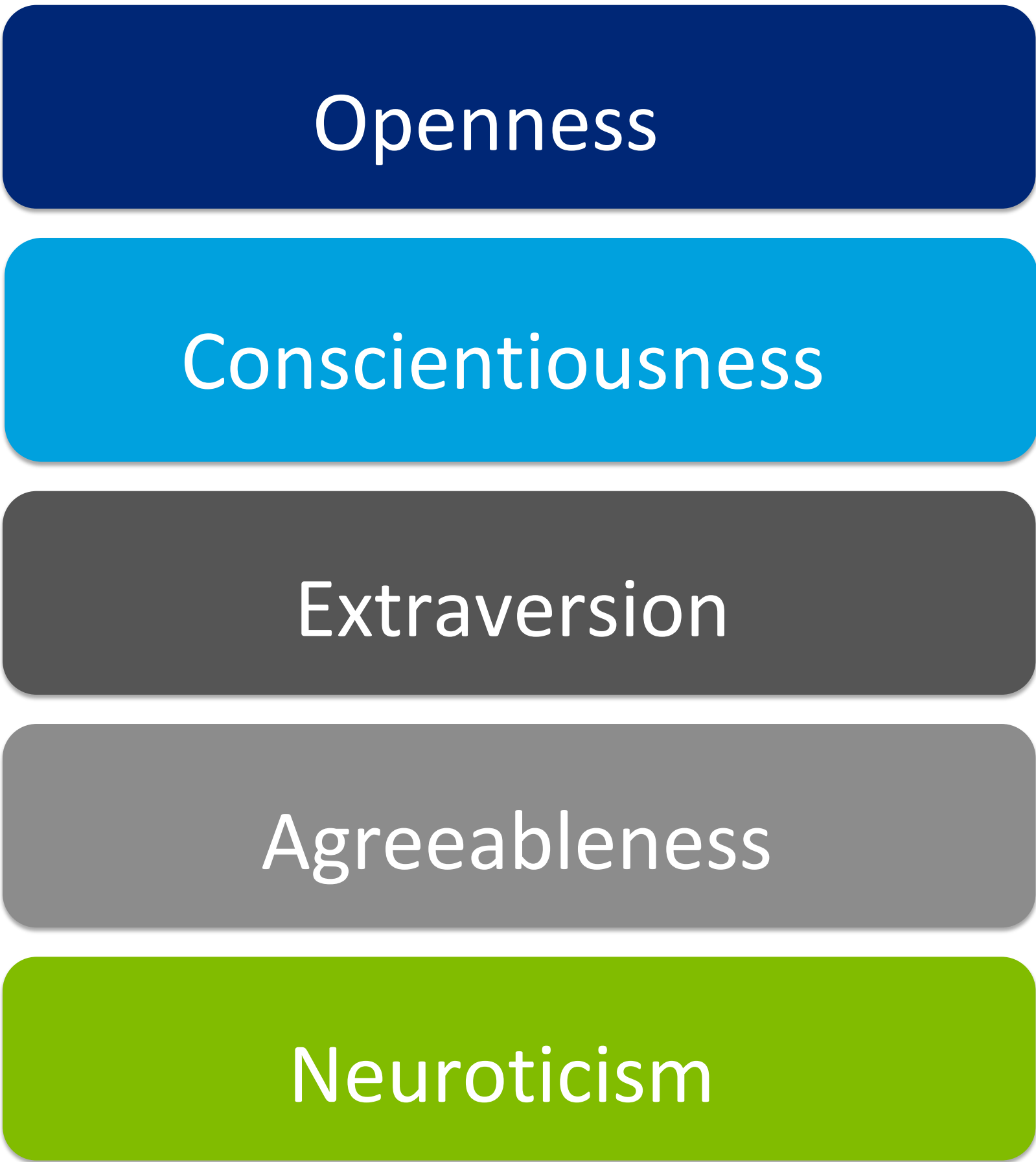
Deloitte & Touche LLP



1). Define: Leverage academic sources to derive taxonomy for OCEAN model and identify broad topics to build the sampling frames.

O.C.E.A.N Model

OCEAN is the acronym for the five-factor model, also known as the Big Five personality types.
-McCrae & Costa (1987)



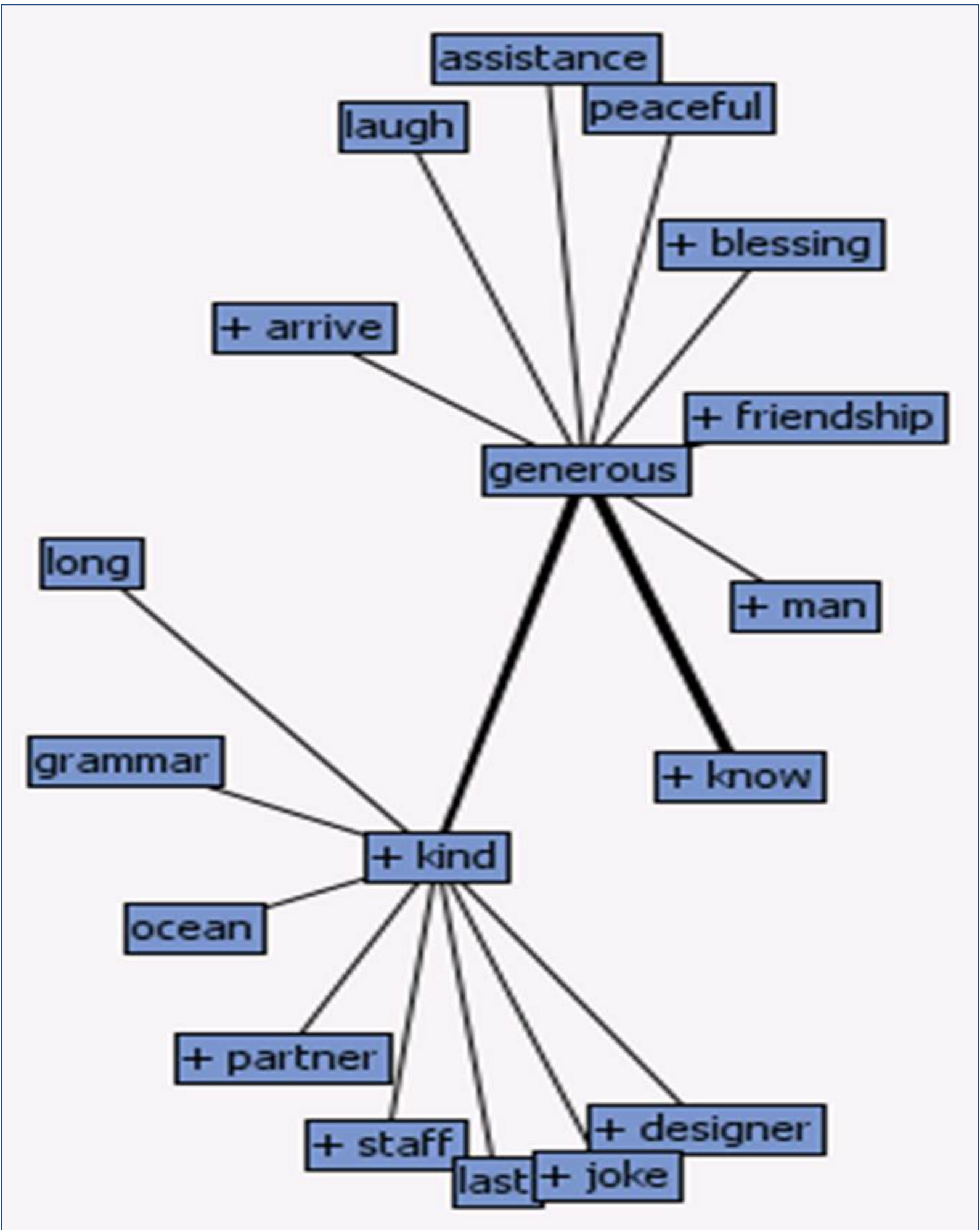
Personality Type Taxonomy

After collecting recognized literature on the subject, we were able to develop a taxonomy of weighted terms used to define the five personality types.



Concept Linking

Concept linking is a way to find and display the terms that are highly associated with another term. The selected term is surrounded by terms that correlate strongest with one another.



Factors Influencing Personality Types

We identified these seven lifestyle focus areas that were subjects throughout our data set.



Religion



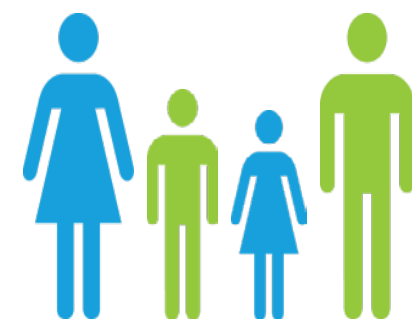
Education



Art



Family



Food



Military Service



Career



Discovering Personality Type through Text Mining in SAS Enterprise Miner 12.1

Deloitte & Touche LLP

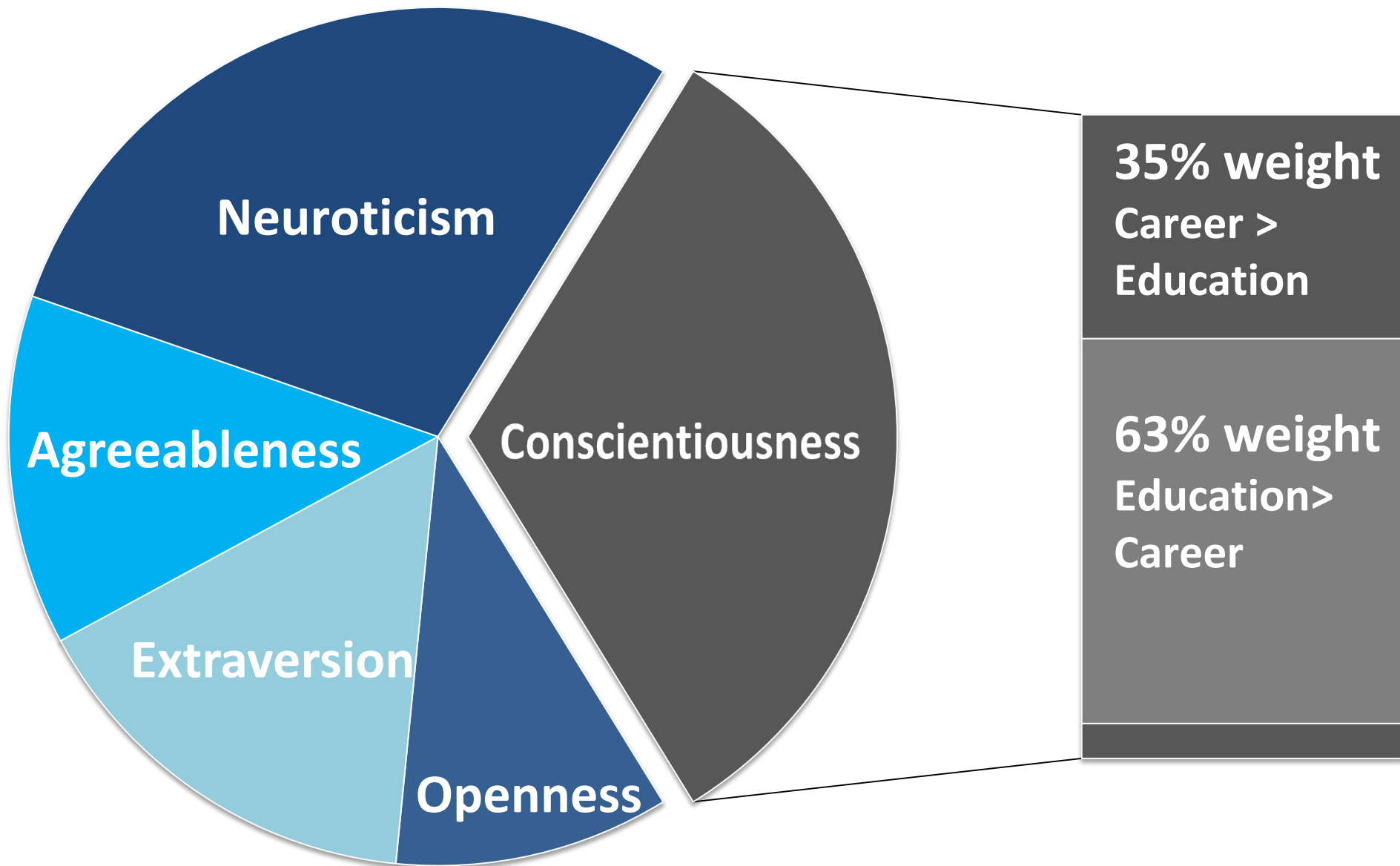


2). Measure: Quantify intangible characteristics by building target based term weights and assess relative importance of intangible characteristics for each particular individual.

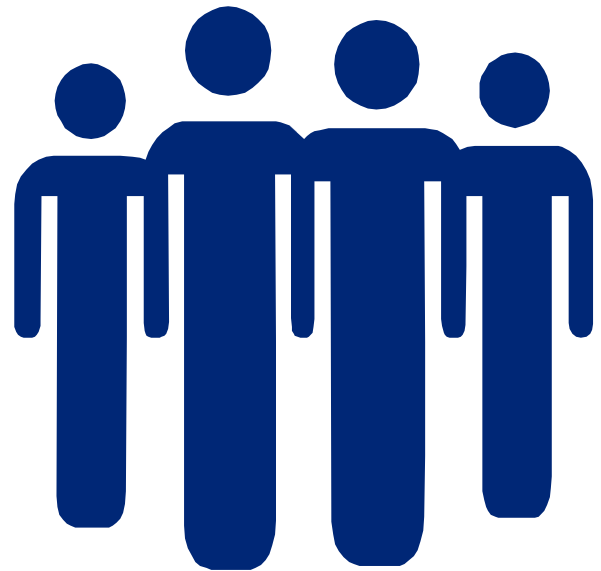
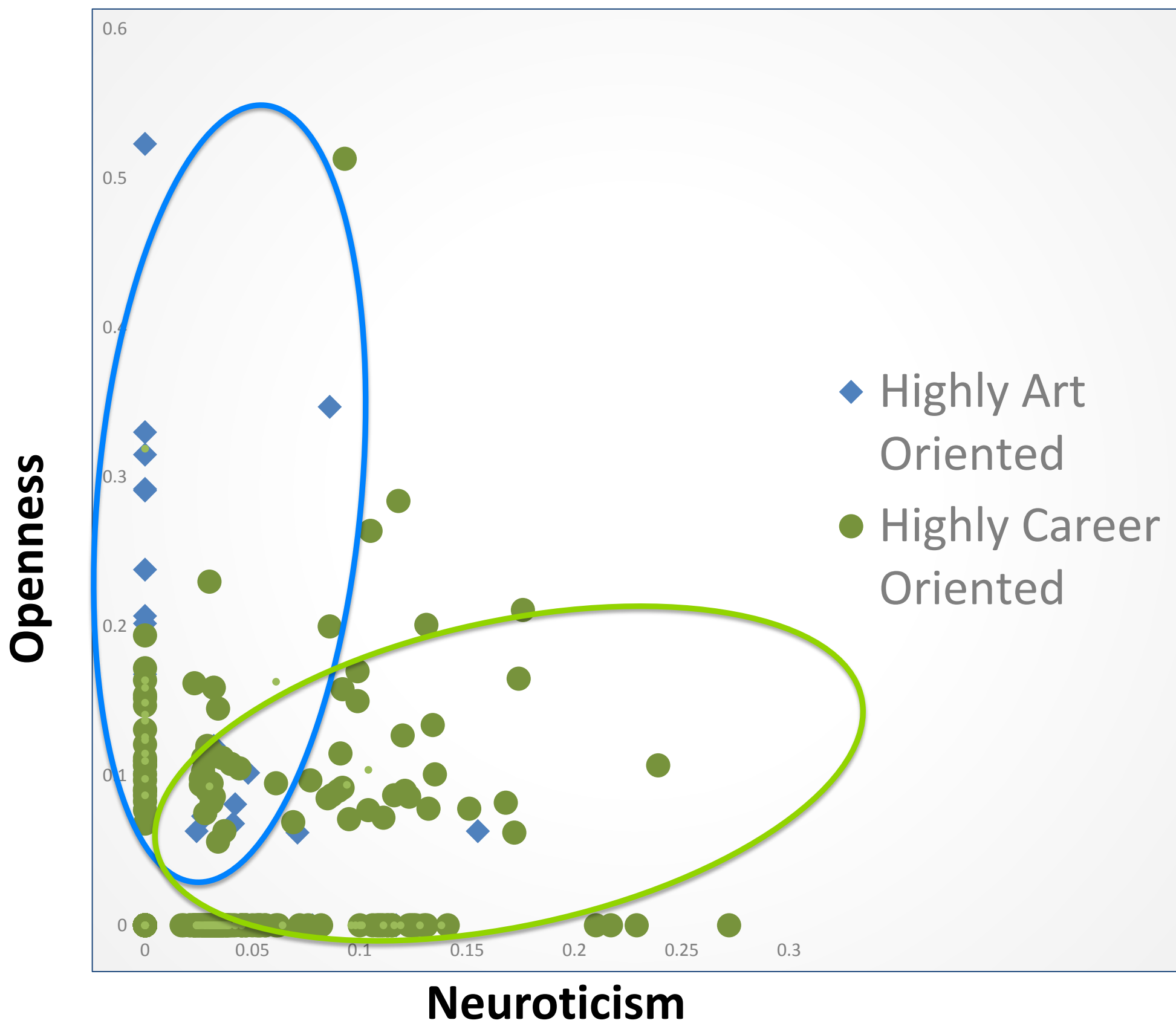
Measuring Personality Types

Each of the five personality traits have been qualitatively described so that they can be quantitatively measured.

A subset of the observations having conscientiousness as the strongest personality trait is dissected to explore two heavily correlated orientations: career and education.



Scatterplot Showing Openness vs. Neuroticism



- Generally speaking, artistic people tend to be more open than neurotic
- Career oriented people tend to be more neurotic than open
- Those with high education orientation tend to be more neurotic and conscientious

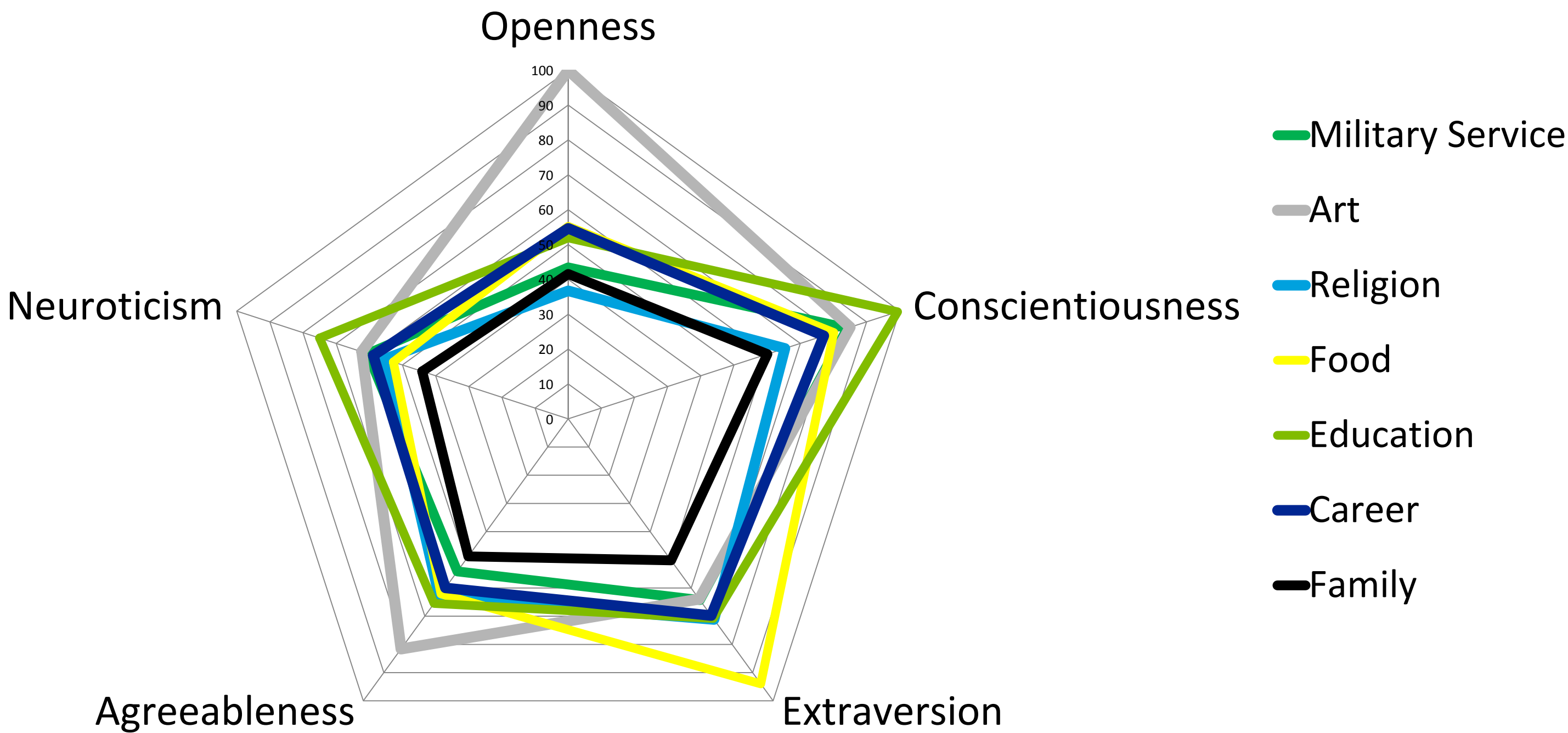
Correlation Matrix: Personality Type vs. Orientation

The heat-map below shows the statistical relationship between personality traits and factors that influence personality.

	Arts	Career	Education	Family	Food	Military Service	Religion
Openness	0.4378449	0.1771027	0.1680627	-0.040112	0.0592454	0.0108119	-0.047497
Conscientiousness	0.0498628	0.3135884	0.3364557	0.1128375	0.0404937	0.1127446	0.0270393
Extraversion	0.0475115	0.3131627	0.210829	0.090644	0.1281069	0.0598953	-0.010933
Agreeableness	0.1213941	0.1627867	0.1537845	0.0699721	0.0215567	0.0089041	0.0744971
Neuroticism	0.0433336	0.2637758	0.334497	0.12286	0.0272014	0.0846723	0.0940988

Normalized Personality Scores

- Personality scores are derived by normalizing the average topic weights for each segment of the population
- Orientations are not mutually exclusive (one individual can be both Career oriented and Family oriented)



Discovering Personality Type through Text Mining in SAS Enterprise Miner 12.1

Deloitte & Touche LLP



3). Analyze: Evaluate statistical relationships between personality traits and classifications and employ multivariate models to predict the magnitude of a personality.

Model Selection Criteria

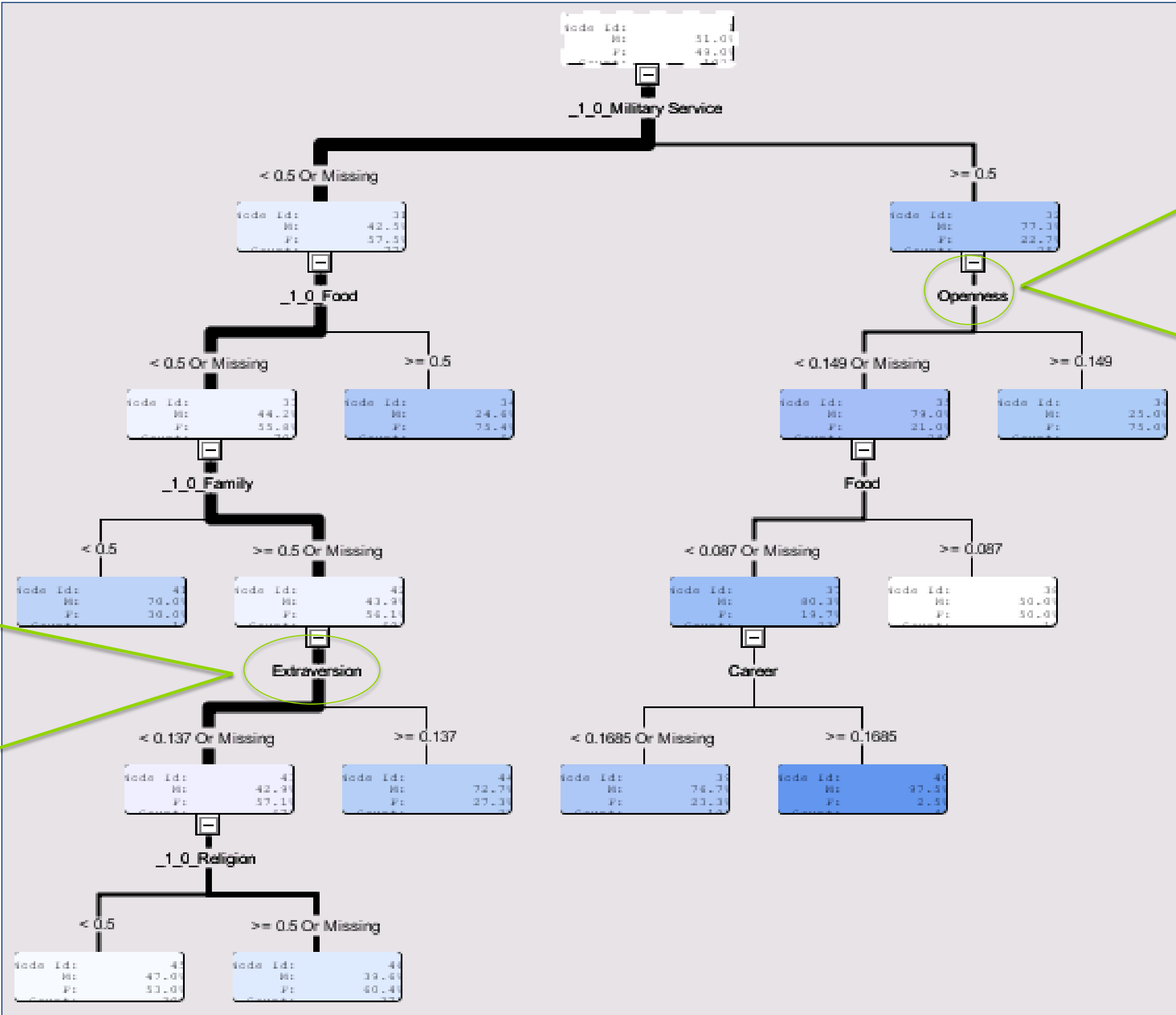
After running this analysis in SAS Enterprise Miner, we were able to select the decision tree as our multivariate model.

Multivariate Model	ROC Index	Lift	Misclassification Rate
Probit Regression	0.623	1.508989	0.417969
Decision Tree	0.656	1.583278	0.349609

Decision Tree

Decision tree model was selected using the above fit statistics while allowing for acceptable interpretability from business users. Each branch shows how personality and lifestyle inputs influence gender distribution.

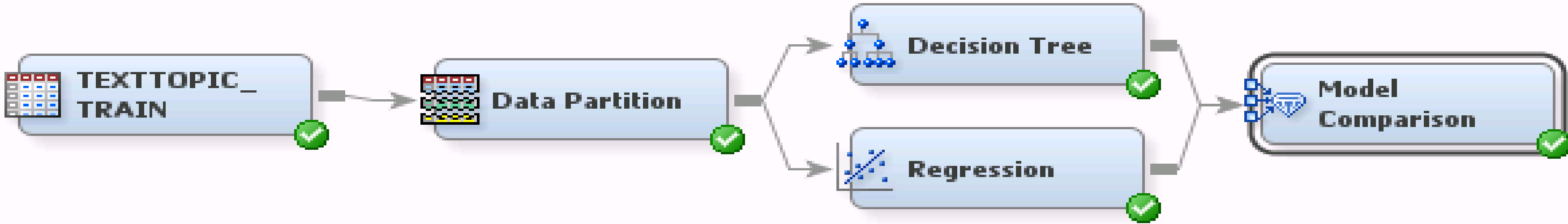
If an individual is characterized as not in the military, food oriented, and family focused he/she is more likely characterized as extraverted.



If an individual is characterized as female and in the military, she would more likely be characterized as open.

SAS Multivariate Model Assessment Node

This node is used to compare various models and the explanatory power of those models.



Scorecard of an Individual

Human capital deployment can be improved by applying risk based scorecards to seek out talent, improve work-group efficiency and develop a diverse and competitive corporate culture.

Personality Scorecard: Individual Match

Name: Jon Doe
Age: 65
Gender: Male
Education: High
Artistic: No
Military: Yes

	Career	Family	Food	Religion
Openness	0.1771027	-0.040112	0.0592454	-0.047497
Conscientiousness	0.3135884	0.1128375	0.0404937	0.0270393
Extraversion	0.3131627	0.090644	0.1281069	-0.010933
Agreeableness	0.1627867	0.0699721	0.0215567	0.0744971
Neuroticism	0.2637758	0.12286	0.0272014	0.0940988

Trait	Career	Family	Food	Religion
Openness				
Conscientiousness				
Extraversion				
Agreeableness				
Neuroticism				



This presentation contains general information only and Deloitte is not, by means of this presentation, rendering accounting, business, financial, investment, legal, tax, or other professional advice or services. This presentation is not a substitute for such professional advice or services, nor should it be used as a basis for any decision or action that may affect your business. Before making any decision or taking any action that may affect your business, you should consult a qualified professional advisor.

Deloitte shall not be responsible for any loss sustained by any person who relies on this presentation.

About Deloitte

Deloitte refers to one or more of Deloitte Touche Tohmatsu Limited, a UK private company limited by guarantee (“DTTL”), its network of member firms, and their related entities. DTTL and each of its member firms are legally separate and independent entities. DTTL (also referred to as “Deloitte Global”) does not provide services to clients. Please see www.deloitte.com/about for a detailed description of DTTL and its member firms. Please see www.deloitte.com/us/about for a detailed description of the legal structure of Deloitte LLP and its subsidiaries. Certain services may not be available to attest clients under the rules and regulations of public accounting.



April 26-29
Dallas, TX

