CARFAX
Banking & Insurance Group
MORE INFORMATION. BETTER DECISIONS.

PINNACLE
ACTUARIAL RESOURCES, INC.

GLMS, MACHINE LEARNING, & MORE, OH MY!

Michael Chen, Pinnacle Actuarial Resources, Inc.,
Gary Wang, Pinnacle Actuarial Resources, Inc.,
Don Hendriks, CARFAX

2018 CAS ANNUAL MEETING – LAS VEGAS, NV

About the Presenters

- **Donald Hendriks**, ACAS, MAAA
- CARFAX
- National Business Consultant
- Greater Detroit Area, MI

- **Michael Chen**, FCAS, MAAA, CSPA
- Pinnacle Actuarial Resources, Inc.
- Consulting Actuary
- Des Moines, IA

- **Gary Wang**, FCAS, MAAA, CSPA
- Pinnacle Actuarial Resources, Inc.
- Senior Consulting Actuary
- Bloomington, IL

CARFAX

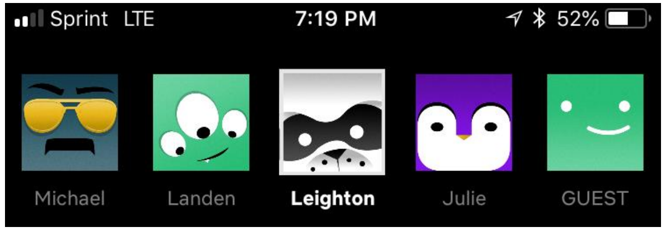
PINNACLE ACTUARIAL RESOURCES, INC.

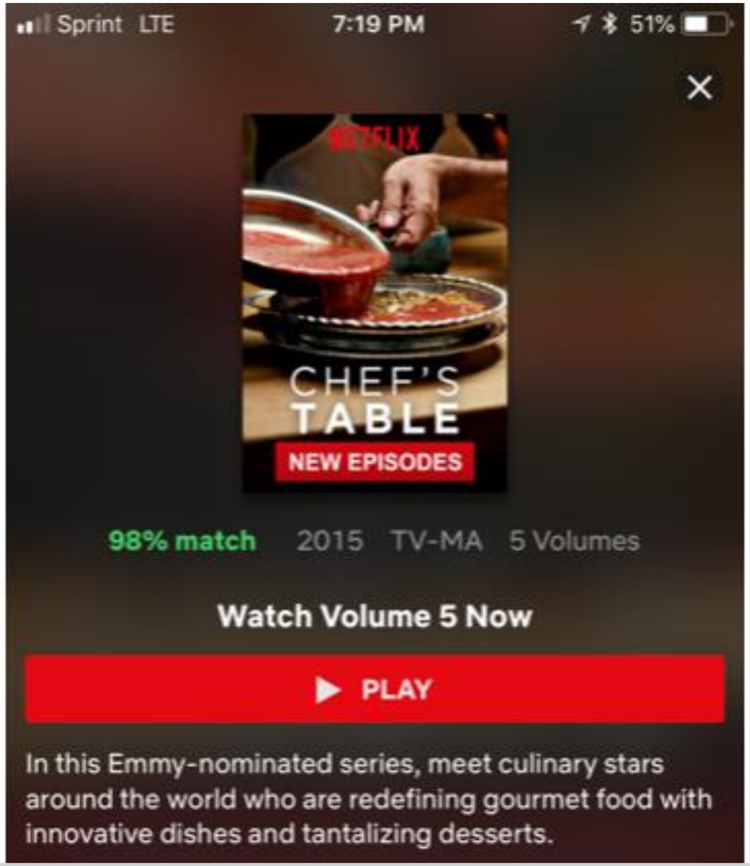
Agenda

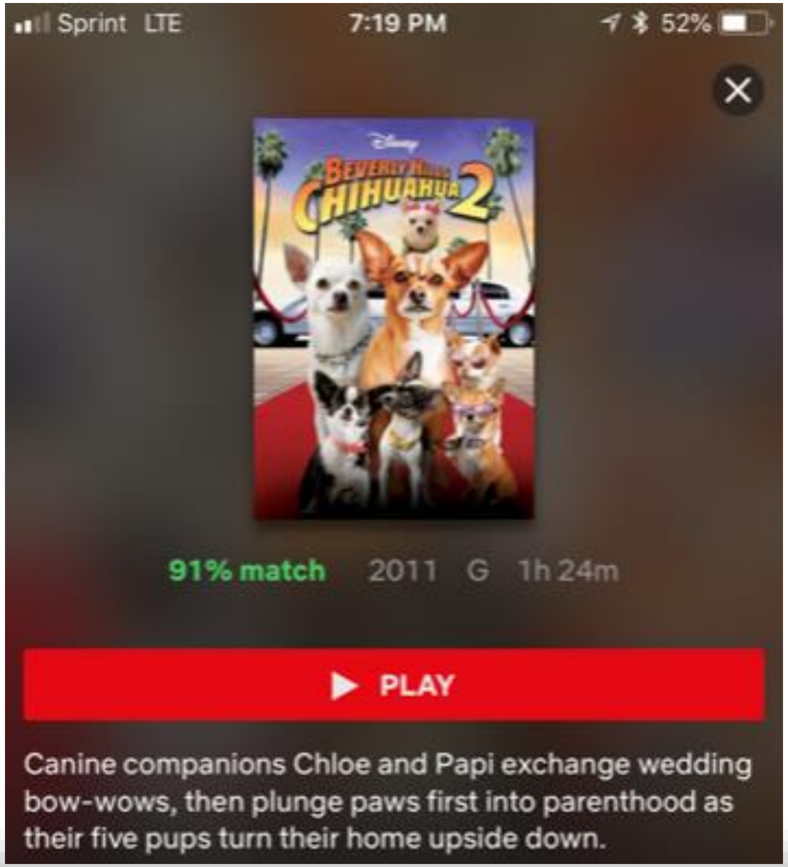
- Introduction
- Residual Analysis – Feature Engineering
- Machine Learning
- Understanding through Competing Models

Introduction

Netflix Prize



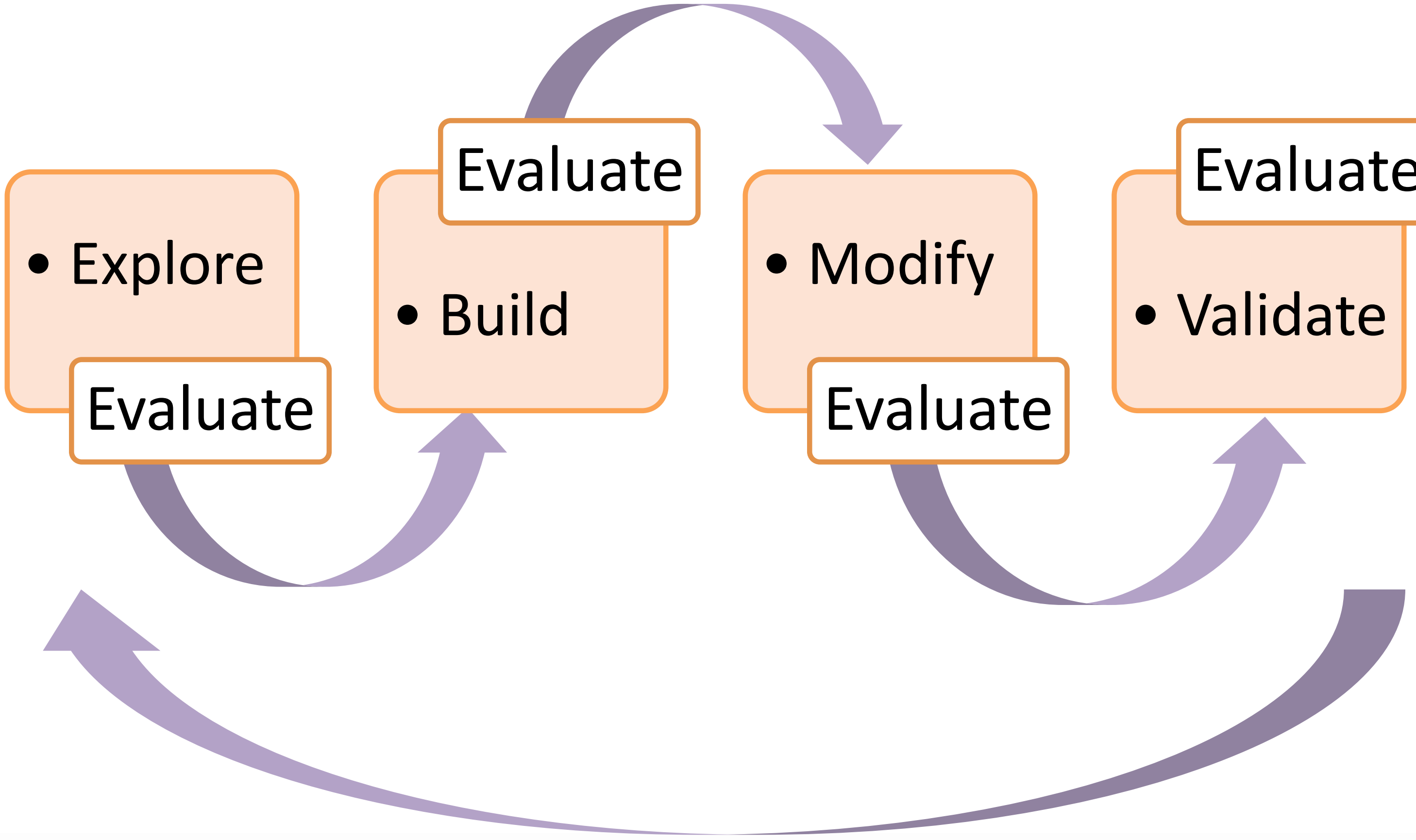




Netflix Prize

- Prize for best collaborative filtering algorithm to predict user ratings for films
- Competition began 10/2/2006 for Netflix, an online DVD-rental and video streaming service
- $1,000,000 awarded in 9/21/2009 to BellKor's Pragmatic Chaos which bested Netflix's own algorithm for predicting ratings by 10.06%

Modeling Process Overview



Feature Engineering

- Feature engineering, refers to the process of creating new input features.
  - Feature engineering is an effective method of improving predictive models.

Feature Engineering, creating useful features

- Calculate statistics like the minimums, maximums, averages, medians and ranges.
  - Investigating the extremes (or the lack) may help define interesting behaviors.

- Create flags and count occurrences of events, highlighting statistically interesting habitual behaviors.
  - NSF notice on renewals for retention analysis
  - Examples: Younger drivers may separate themselves more clearly in a Non-Standard book than older drivers

- Create ratios seeking to add predictive value to already meaningful variables.
  - density, population/land area
  - vehicle to driver ratio (often used in a Matrix)

Feature Engineering, creating useful features

- Develop quintiles across variables of interest seeking to create expressive segments of the population while also dealing with extreme values.
  - Creating bins to transform variates to categorical variables

- Apply dimensionality reduction techniques, ranks, clustering etc. expecting that grouping those with similar behaviors will be statistically beneficial.
  - Principal Components
  - Clustering

Feature Engineering, creating useful features

- Consider the element of time as an important interaction with any feature that has been developed.
  - Recognizing newer policies are biased toward no claim history, opportunity to tap into external data

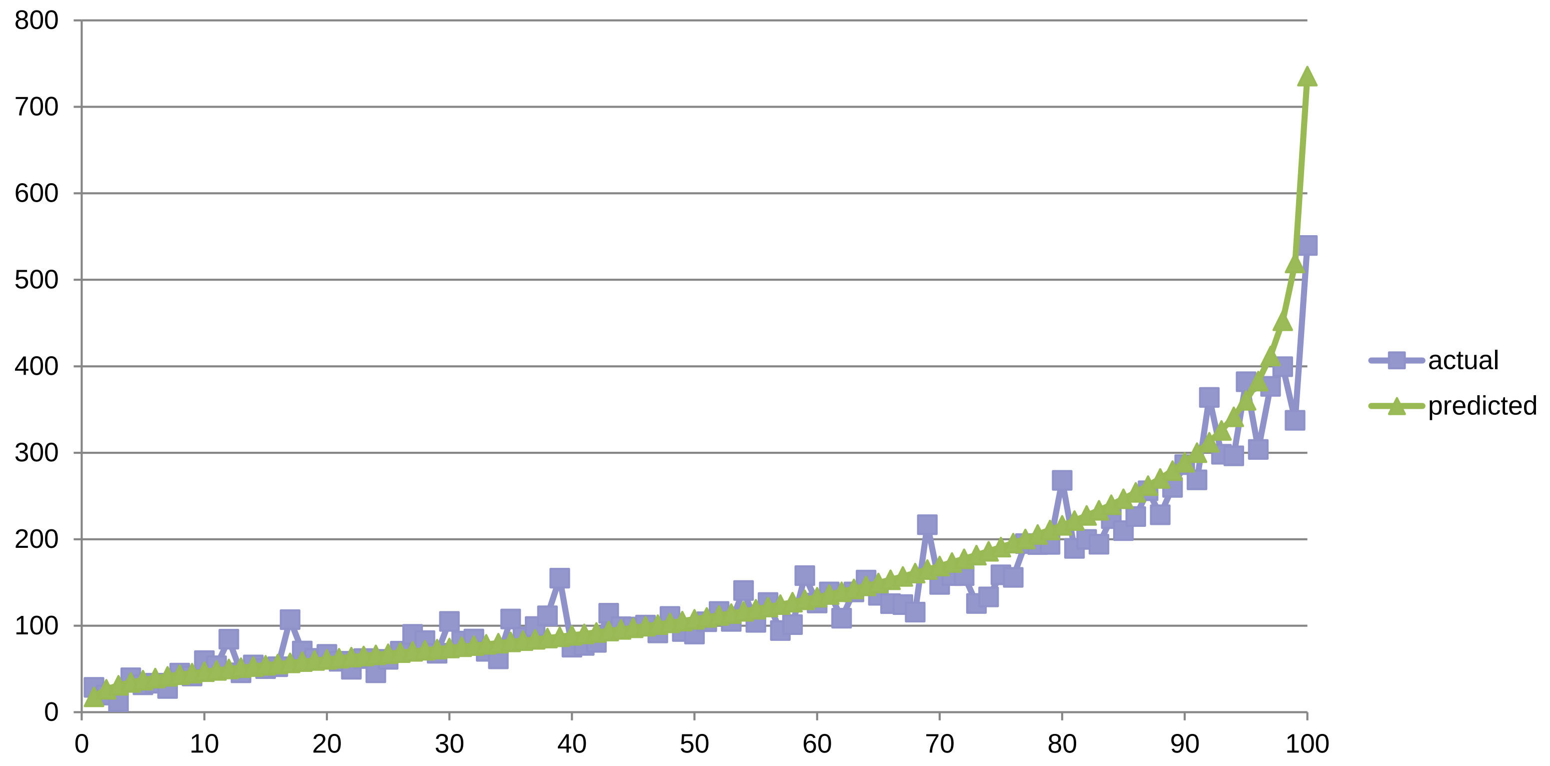
- Use regression to identify trends in continuous variables thinking that moves up or down (whether fast or slow) will be interesting.
  - Looking at univariate and doing a fit

Feature Engineering, using non-linear techniques

- Have GLM
- Create residuals from you Actual and Predicted Values
- Model on the residuals using non-linear techniques
- Practical considerations concerning algorithm from the models
- Score your original GLM dataset
- Model with new variable either as a variate or could group as a categorical
  - Modeling techniques like decision trees naturally produce bins (categorical), bins have values so could also be treated as a variate
- Alternatively could do similar exercise on the actual for example vehicle symboling

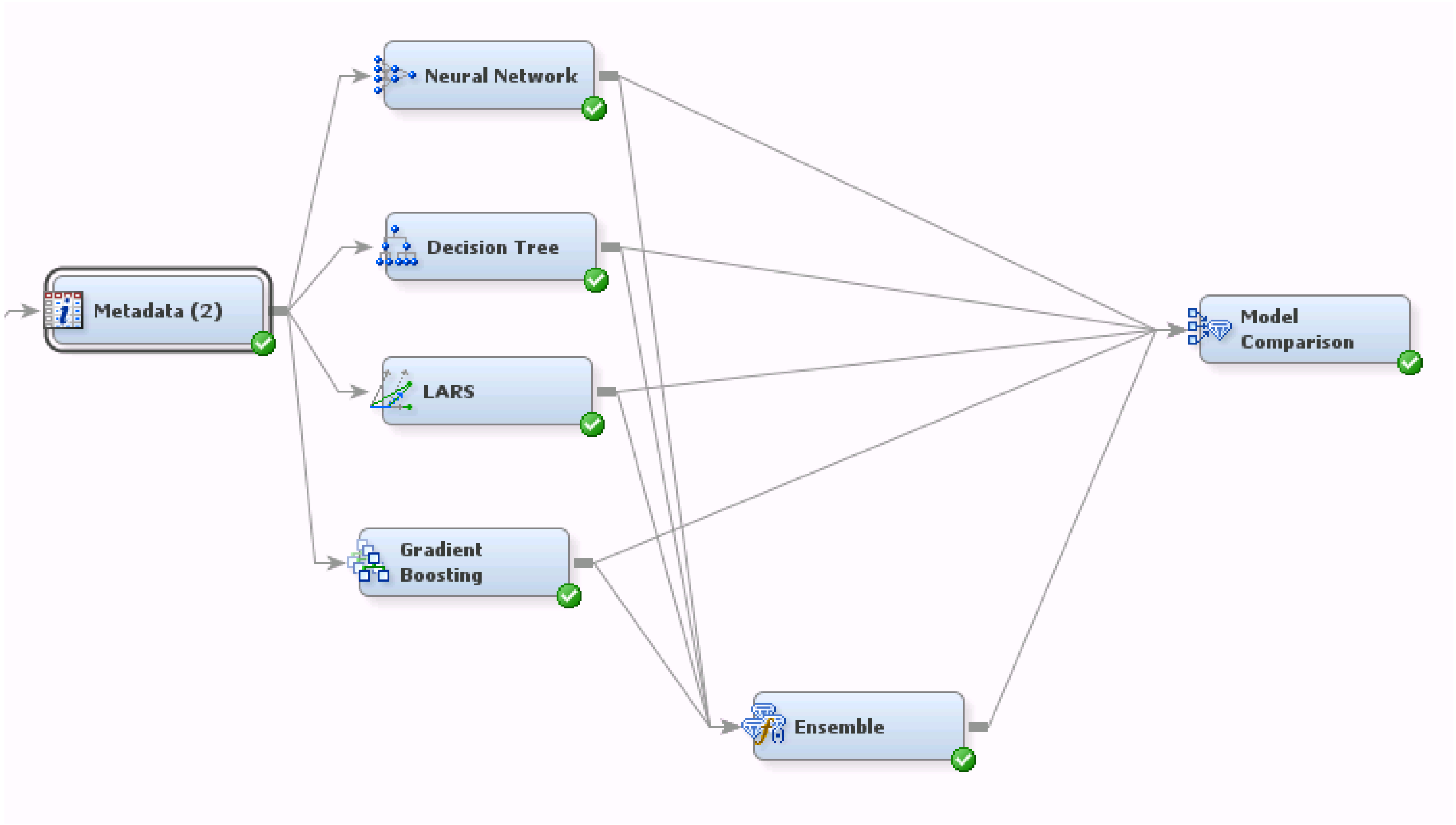
Feature Engineering, using non-linear techniques

- Initial GLM has been completed
  - Validates fairly well but may have room for improvement



Feature Engineering, using non-linear techniques

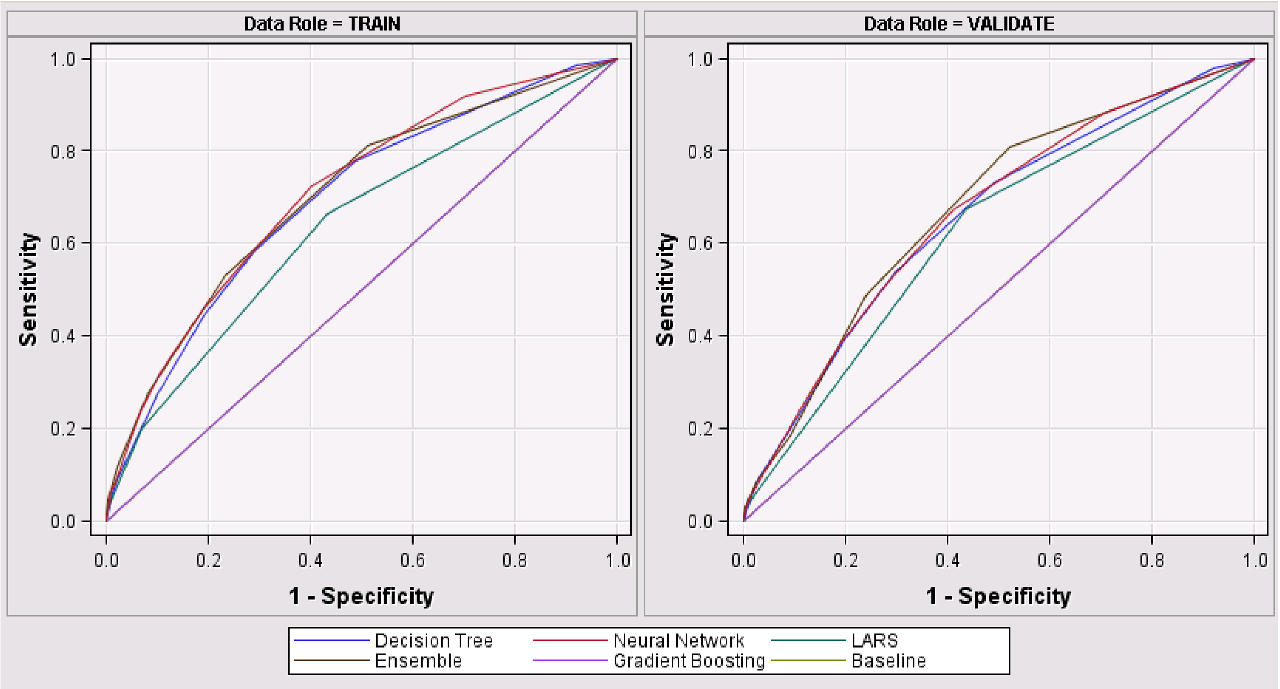
- Model on the residuals using non-linear techniques



Feature Engineering, using non-linear techniques

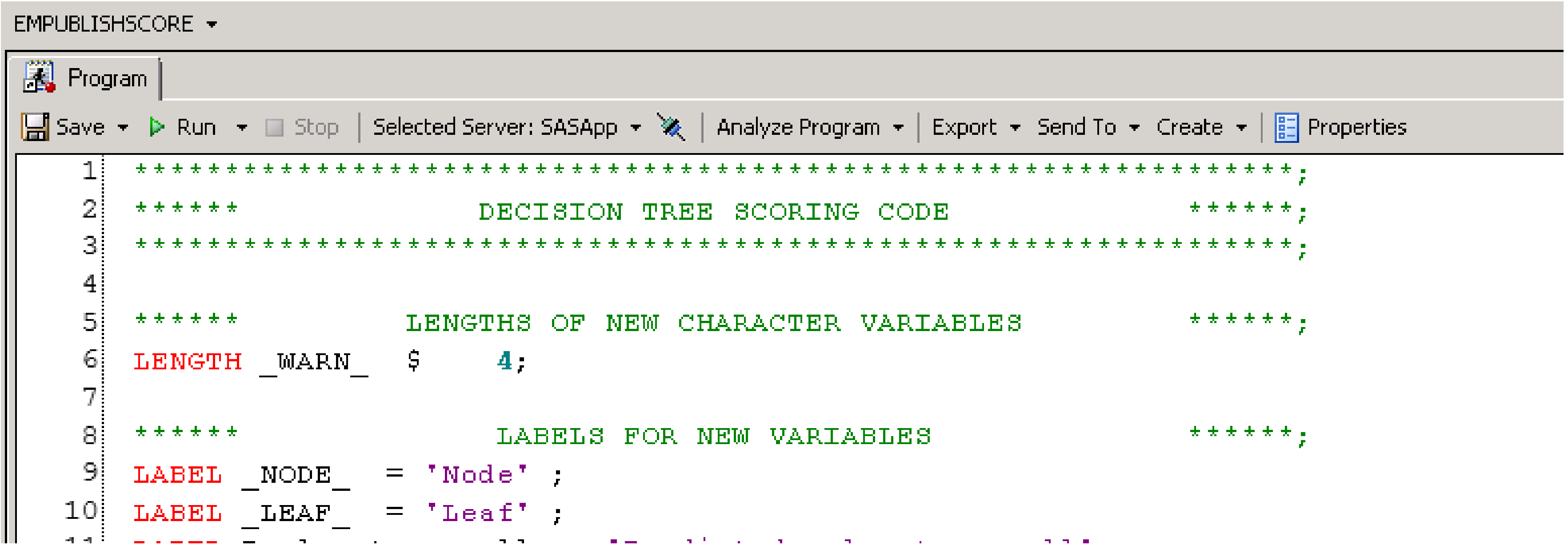
- Practical considerations concerning algorithm from the models

Selected Model	Model Description
Y	Ensemble
	Neural Network
	Decision Tree
	Gradient Boosting
	LARS



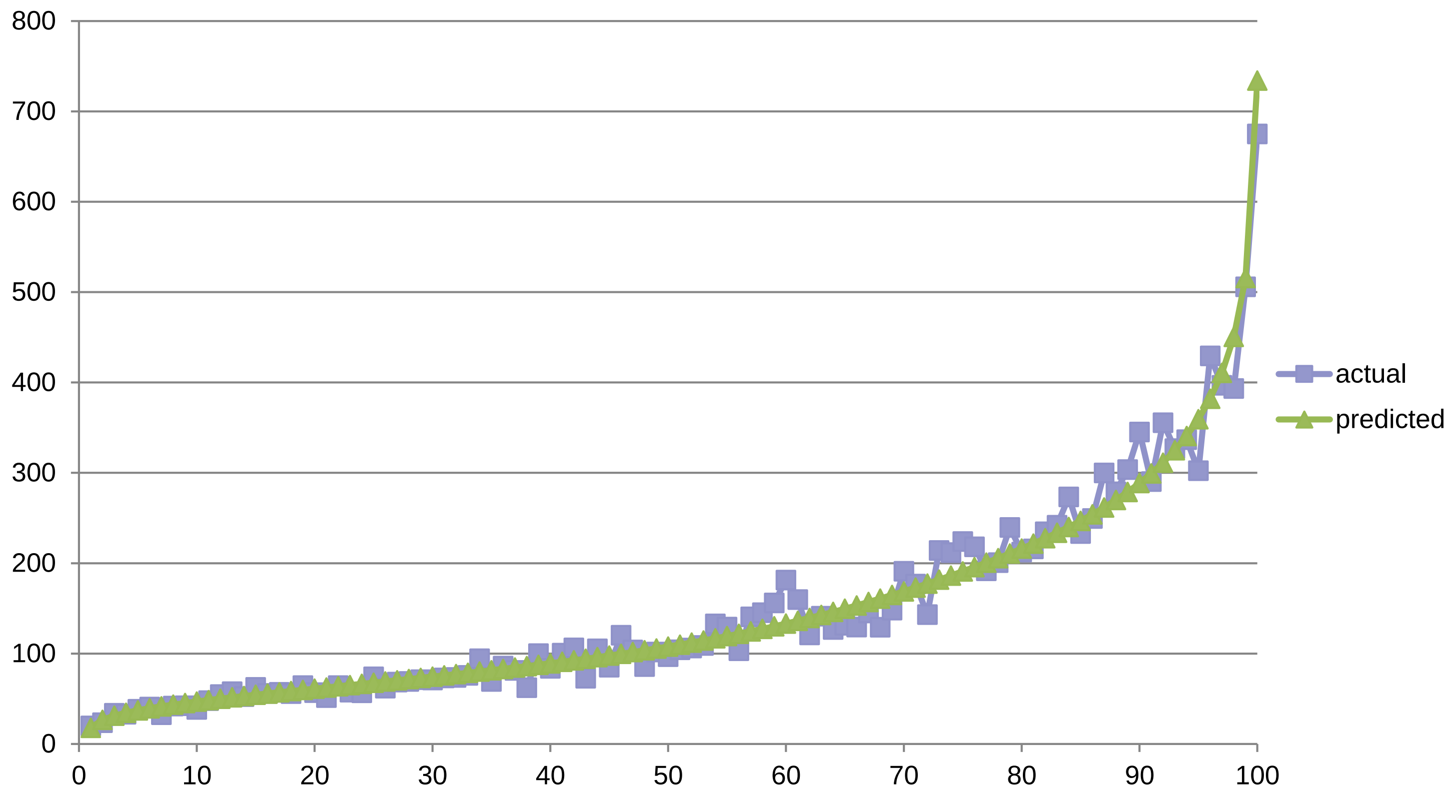
Feature Engineering, using non-linear techniques

- Use developed non-linear model to score your original GLM dataset



Feature Engineering, using non-linear techniques

- Model with new variable either as a variate or could group as a categorical



Questions

Thank You for Your Time and Attention

Don Hendriks

donhendriks@carfax.com

Michael Chen

mchen@pinnacleactuaries.com

Gary Wang

gwang@pinnacleactuaries.com

CARFAX

PINNACLE
ACTUARIAL RESOURCES, INC.