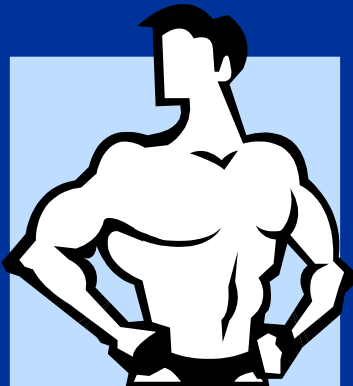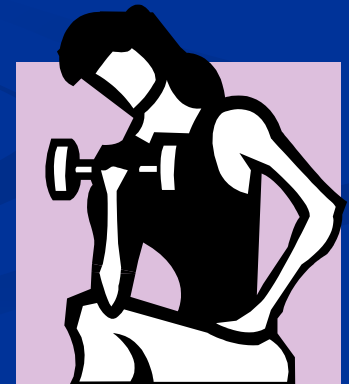# Predicting New Customer Retention for Online Dieting & & Fitness Programs

Team 4

Amy Brunner • Harin Sandhoo

Lilah Pomerance • Paola Nasser

Srinath Bala

Problem

- Gymamerica.com: online online personal trainer

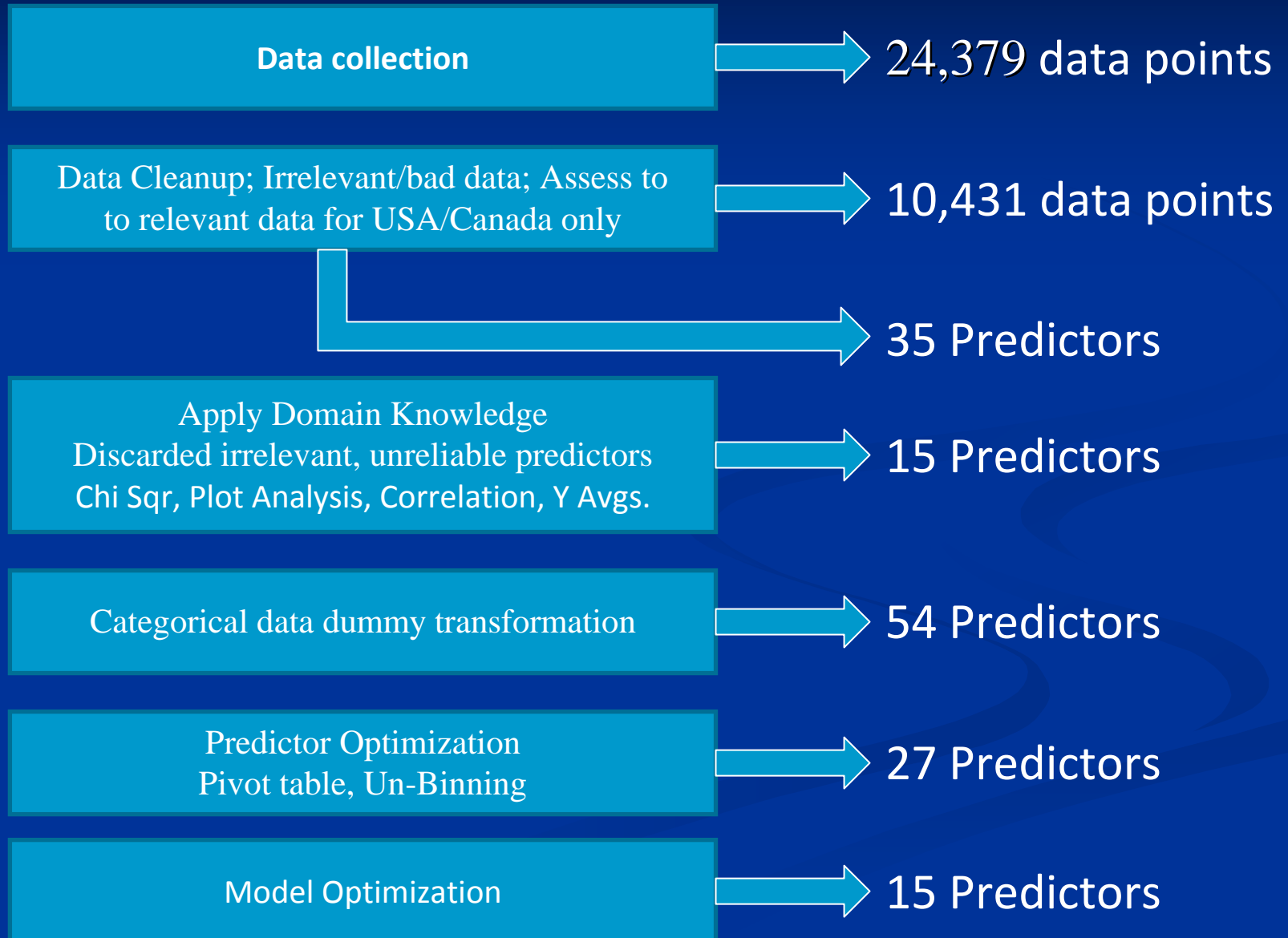- Predict whether a new user user will become a paying paying customer after their their 10-day free trial.

Summary of Data

- Source: Genesant Technologies, Inc.
- Initial Data Set Size – 24,379 observations and 37 variables
- Cleaned up data – 10,431 observations
  - US/Canada customers with exercise application that registered between 1/1/04 and and 6/30/07.
  - Excludes: Test customers, Failed CC, customer who failed to finish their exercise and diet and diet configuration.
  - Numerical & Categorical
- Defined variables and formats
  - Output variable - Billing State (Paid or Cancelled)
  - Predictors
    - User profile: Gender, Weight, Height, Location, SignupDate
    - Strength Specific parameters
    - Cardio Specific parameters
    - Diet Specific parameters
    - Miscellaneous: isSpecialOffer, PackageAppType, PackagePrice

Data Preprocessing Summary

Data collection	→ 24,379 data points
Data Cleanup; Irrelevant/bad data; Assess to to relevant data for USA/Canada only	→ 10,431 data points
	→ 35 Predictors
Apply Domain Knowledge Discarded irrelevant, unreliable predictors Chi Sqr, Plot Analysis, Correlation, Y Avgs.	→ 15 Predictors
Categorical data dummy transformation	→ 54 Predictors
Predictor Optimization Pivot table, Un-Binning	→ 27 Predictors
Model Optimization	→ 15 Predictors

Exploration

workoutPlan	count	avg
Advanced 5K w/ Weight Training	33	0.6363636
Advanced Toning for Men	381	0.6351706
Advanced Power Workout	427	0.5971897
Select Cardio Only Workout Plan	113	0.5044248
21-Day Advanced Core Toning	169	0.6153846
Advanced Circuit Training	275	0.5527273
Advanced Toning for Women	340	0.6970588
Select Workout Plan	8693	0.5857587

strWktLength	count	avg
30 minutes	4383	0.5496235
45 minutes	169	0.6153846
0 minutes	113	0.5044248
120 minutes	166	0.6144578
60 minutes	4814	0.6208974
90 minutes	786	0.6348601

dietPlan	count	avg
Food Pyramid Diet	1044	0.4846743
High Protein Diet	3239	0.5637542
NO DIET	3481	0.6595806
Macronutrient Diet	2667	0.5744282

weightGoal	count	avg
Lose Weight	5800	0.5477586
Maintain Weight	294	0.5408163
NO DIET	3481	0.6595806
Gain Weight	856	0.6168224

dietConveniencePlan	count	avg
The On-The-Go Meal Plan	2312	0.5977509
The Combination Meal Plan	2532	0.5533175
The Home Cooking Meal Plan	443	0.5553047
The Easy Home Meal Plan	1659	0.5009042
NO DIET	3485	0.6599713

Modeling

	Training Set Error%	Validation Set Error%	Test Set Error%
Logistic Regression	37.33%	39.76%	37.20%
KNN	35.01%	41.48%	37.44%
Classification Tree	38.90%	41.58%	38.21%
Naïve Bayes	38.88%	40.91%	38.83%
Discriminant Analysis	40.43%	42.22%	40.17%
Naïve Rule	40.95%		

	Test Set Sensitivity	Test Set Specificity
Logistic Regression	86.24%	28.23%
KNN	85.68%	28.47%
Classification Tree	89.38%	18.62%
Naïve Bayes	78.12%	36.18%
Discriminant Analysis	63.64%	54.21%

Best Model : Logistic Regression

■ Better at correctly detecting paying customers (Sensitivity = 86.24%)

■ Ability to rule them is not so good (Specificity = 28.23%).

Input variables	Coefficient	Std. Error	p-value	Odds
Constant term	-0.24233337	0.48623437	0.61821061	*
gender_M	0.63393086	0.08208836	0	1.88500571
squatPreference_yes	0.23397914	0.06712396	0.00049071	1.26361811
log(strAbility)	-0.23567867	0.06456982	0.00026226	0.79003447
log(age)	0.14171952	0.10306153	0.16910163	1.15225339
strWktLoc_atGymFreeWts	0.70793337	0.22595046	0.0017295	2.02979207
strWktLoc_atGymFreeWts&Machines	1.35648978	0.15434419	0	3.8825407
strWktLoc_atGymMachines	1.11714578	0.17428556	0	3.05611873
strWktLoc_atHomeDumbells	1.22130656	0.17483339	0	3.39161611
strWktLoc_atHomeFreeWts	0.80089557	0.18768552	0.00001979	2.22753501
strWktLoc_atHomeFreeWts&LegExtMachine	1.31087041	0.18496068	0	3.70940113
strWktLen_45 minutes or more	0.10658011	0.06212238	0.08622657	1.11246705
wktPlan_Advanced Toning for Women	0.70400673	0.20847158	0.00073283	2.02183747
wktPlan_Select Workout Plan	-0.61317694	0.11918895	0.00000027	0.54162741
dietPlan_NO DIET	0.31598803	0.06445931	0.00000095	1.37161386
dietPlan_PyramidDiet	-0.11608823	0.09830552	0.23764549	0.89039665

Test Data scoring - Summary Report

Cut off Prob.Val. for Success (Updatable)	0.5

Classification Confusion Matrix		
	Predicted Class	
Actual Class	payed	cancelled
payed	1072	171
cancelled	605	238

Error Report			
Class	# Cases	# Errors	% Error
payed	1243	171	13.76
cancelled	843	605	71.77
Overall	2086	776	37.20

Conclusions: Customer Profile

Male

No Diet Plan Selected

Like doing barbell Squats
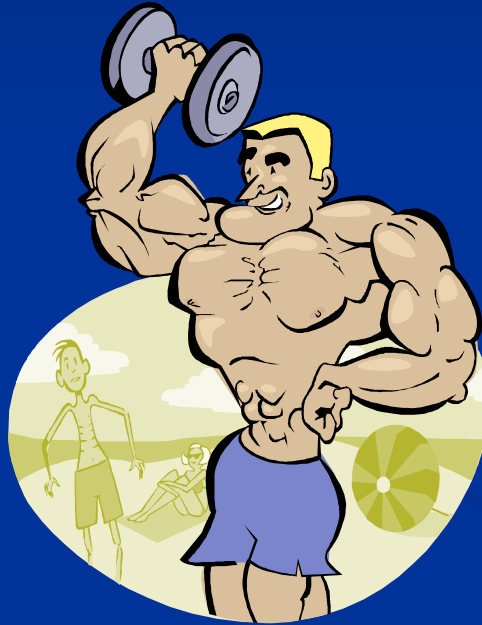
Workout Location: Home or Gym with Free weights and Machinery

*If female, Advanced Toning for Women is significant

Recommendations

- Pursue a marketing campaign targeted towards individuals who fit the previous previous profile
- Send special offers/incentives for people that use the fitness profile, fail to enter the enter the free trial, but by the model are likely to be a paying customers.
- Change the order of the workout plan options depending on gender, so that female female users see "Advanced Toning for Women" workout at the top and as the the default.
- Improve data storage of user preferences for future more accurate models
  - Age
  - Strength/Cardio Experience
  - Keep history of preferences not just more recent one
- Continue to provide options for working out at home or at the gym
- Improve diet prescriptions for male population and create more offers/incentives for offers/incentives for users to use it.

Date of Birth: December / 11 / 1989

1. **Which of the following best describes your experience with strength training?**
   - ○ I have never trained before
   - ○ I have not trained in years
   - ○ I have trained periodically over the last two years
   - ◉ I have been training on a regular basis for less than six months
   - ○ I have been training on a regular basis for more than six months

QUESTIONS?