

# Analytical Report on Pampered Pets Inc.

- Planned sales approach: PPI wants to expand its sales, using a more targeted sales approach. PPI sales agents will call selected prospective customers with a cost, based on newly available detailed demographic data. The correctly phoned pet owner will bring about profit.
- Description of prediction process
- Evaluation of prediction of potential customers

# Report Goal

- ▶ Evaluate the performance of methods used on predicting targeted phone advertisement.
- ▶ Provide supporting information for decision made on future of the planned sales approach.
- ▶ Evaluate dataset of potential customer properties collected by corporation.
- ▶ Address possible improvements on data manipulation.

# Dataset Exploration

- ▶ 28 demographic attributes, including household type, rent etc.
- ▶ Enough observations to test predictions
- ▶ Pre-process the dataset: Remove blank/invalid data, remove outliers etc.

# Data Exploration

- ▶ Redundant attributes may tarnish model's predictive power by over-fitting the features.
- ▶ Examples of removal:
  - ▶ Height: Customer's height is not related to dog/cat owning
  - ▶ Years in USA: Unrelated to dog/cat owning
  - ▶ Race: Unrelated to dog/cat owning
- ▶ Data attributes with similar property may waste computing power.
- ▶ Examples of combination:
  - ▶ Working Status & Employment Status
  - ▶ House Tenure & Own or Rent House

*Find detailed removal list and reasons in Appendix*

# Necessary Terms for Model Introduction

Accuracy: Ratio of successfully predicted classes in all predictions.

Expected gain: Expected gain based on predicted result.

# Modeling for Prediction

- ▶ Goal: Given collected potential customer dataset, apply statistical/machine learning models to predict whether they are pet owners, cat owner, both or neither. Then evaluate the gain/loss of phone advertisement based on prediction.
- ▶ Steps
  - ▶ Separate given dataset into training and testing set.
  - ▶ Train models on training test, make prediction on test set
  - ▶ Calculate model accuracy and recall
  - ▶ Compute corresponding gain/loss

# Model 1: K-Nearest Neighbors

- ▶ Method: Assign observations into K groups of neighboring points and assign such grouped observations into targeted class
- ▶ Model Calibration: Choose of different number of neighbors, named K

K number	Accuracy	Expected Gain
K=5	0.499	-31905.5
K=20	0.549	-17322.0
K=100	0.580	2724.0

# Model 2: Logistic Regression

- ▶ Method: A statistical model using logistic function to calculate probability of occurrences and classify observations into dog owner, cat owner, both or neither.
- ▶ Model Calibration: Use of different logistic functions, named Kernel.

Kernel Name	Accuracy	Expected Gain
L1 regularization	0.576	-10130.5
L2 regularization	0.579	-8266.5
No regularization	0.574	6466.5



# Model 3: Support Vector Machine

- ▶ Method: Consider the collected dataset as a plain or cube or higher dimension cube, draw lines that divide dataset into different region, each region represent one designated class
- ▶ Model Calibration: Draw the cube differently, named Kernel.

Kernel Name	Accuracy	Expected Gain
Linear	0.585	0.0
Poly	0.575	-4548.0
Rbf	0.584	374.5

# Evaluation of Data and Models

- ▶ **Model Evaluation:** The accuracy of all three models with different calibration does not exceed 0.60.
- ▶ **Data Evaluation:** The dataset quality needs improvement.
- ▶ **Model Evaluation:** Non-pet owners are classified with high accuracy.
- ▶ **Data evaluation:** Non-pet owner data maintains large proportion in dataset.
- ▶ **Model Evaluation:** SVM model with rbf kernel is the most optimal model for predicting designated pet owners, though the net gain is still negative.

# Recommendation

- According to the predictions generated by three models, the maximum net profit could reach 6466.5.
- The cost per phone may be higher than 1 dollar and given other delivered cost, PPI should not move forward with the current approach.

# Discussion on Data and Models

- ▶ The dataset is dominated by non-pet owners, which brings large bias to models for prediction.
- ▶ The relative return on customers correctly classified compared to loss derived from each phone advertisement made is trivial.

# Suggested Improvements

- ▶ Improve quality of dataset regarding all aspects, such as less non-pet owners, more predicative attributes etc.
- ▶ Put more prediction weight on attributes with more importance for models
- ▶ Implement more algorithms on the dataset.

# Conclusion

- ▶ Models implemented does a good job in predicting, however, failure to consider cost in models weaken their ability to maximize profit.
- ▶ Problems related to dataset and the sensitivity of proposed approach to current situation can be solved to improve prediction power and expected gain.

Thank you

# Appendix

## Reflection

A concise report needs only the demonstration of methods and results rather than detailed description. It also needs to consider the academic background of audience, including necessary background introduction, minimal of texts and essential images.

# Appendix

- ▶ AK23(Home Type)-The home type is not related to CLASS while I think the number of rooms will be more useful. Number of rooms represent the possible space for pets but home type does not.
- ▶ AK25(Own or Rent)-As it is not hard for the owner to travel or move with the pets, whether the household is rented or owned will not affect one's willingness to own a pet.
- ▶ CNTRY5(Country of Birth): There is no relation between owning a pet and owner's birth place.
- ▶ HGHTL\_P(Height): There is no relation between owning a pet and owner's height.
- ▶ RACECEN(Race): There is no relation between owning a pet and owner's race. Pet owners vary a lot in racies around the world.
- ▶ WGHTP\_P(Weight): There is no relation between owning a pet and owner's weight.
- ▶ WLKANY(Walking): There is no relation between owning a pet and owner's walking habit.
- ▶ YRUS(Years in USA): There is no relation between owning a pet and owner's years lived in the USA.