Predicting Personalized Privacy Preferences in the Smart Home

Natã M. Barbosa, Joon S. Park, Yaxing Yao, Yang Wang

1. Problem

Information privacy has been a major subject of discussion around adoption of smart home devices. Privacy can be highly contextual and therefore difficult to approach. While some may be comfortable with collection and secondary use of personal data, others may feel uneasy about such practices.

Therefore, effective privacy-protecting mechanisms must not rely on one-size-fits-all approaches. Instead, such mechanisms should consider each individual's concerns and attitudes, along with contextual factors that can influence decisions about privacy.

This is precisely what the models presented here try to predict.

2. Goals

1. Predict personalized allow/deny decisions.

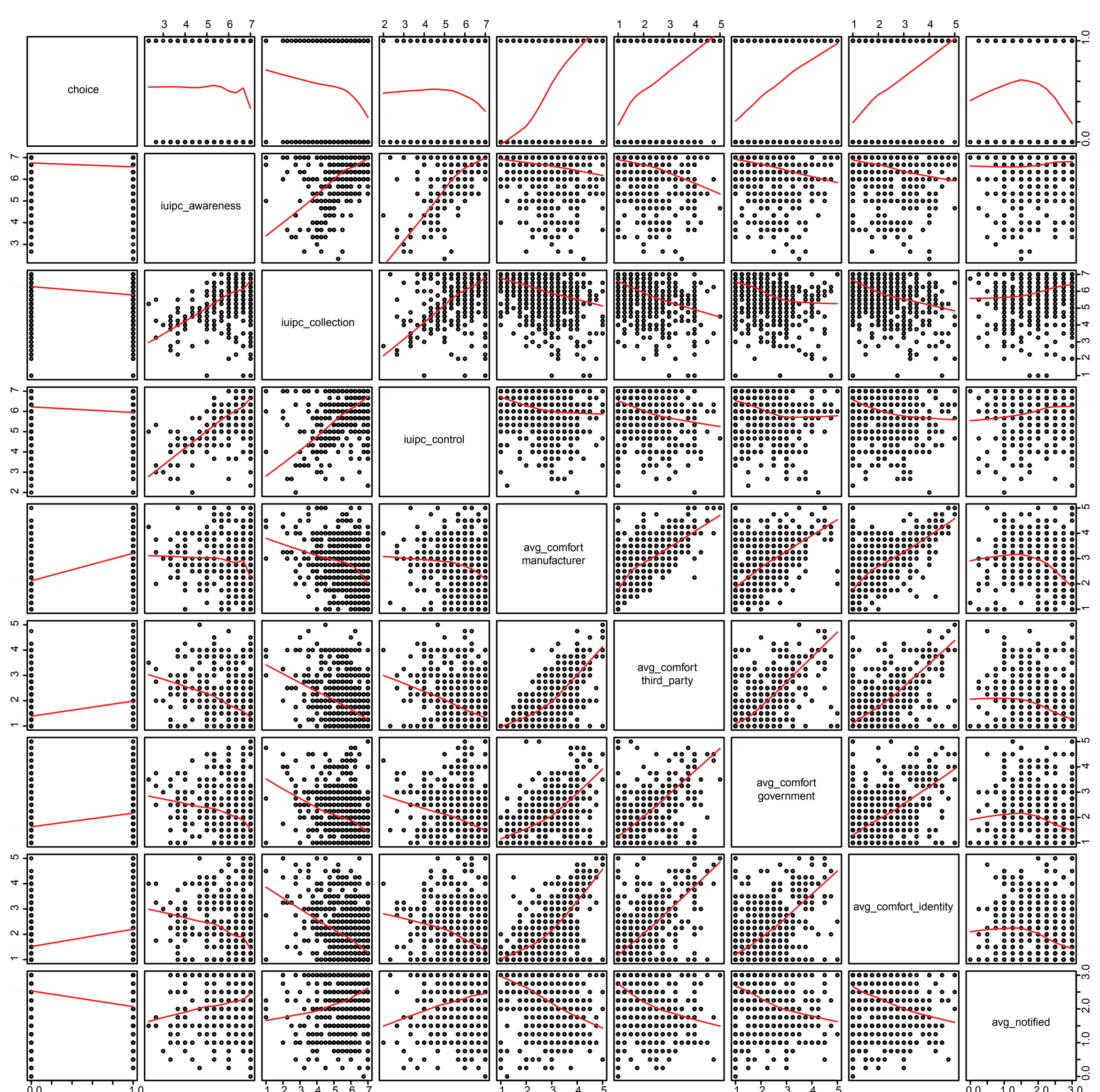
2. Predict situations that could make users less or more comfortable.

3. Predict dollar amount associated with privacy.

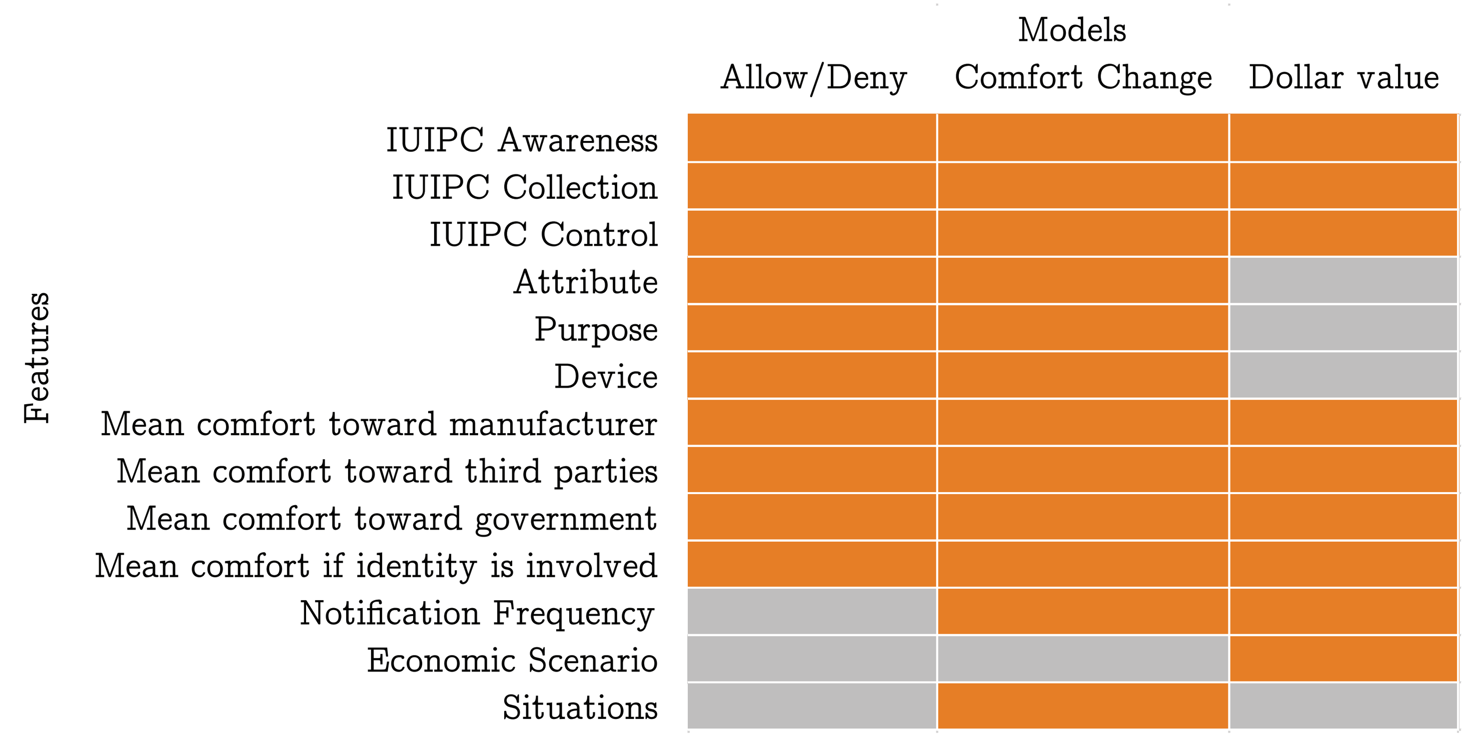
3. Data

I used a dataset collected via an online survey on Amazon Mechanical Turk containing 2,792 combinatorial scenarios involving different attributes, purposes and devices.

Choice *vs.* Numerical Features



4. Models and Features

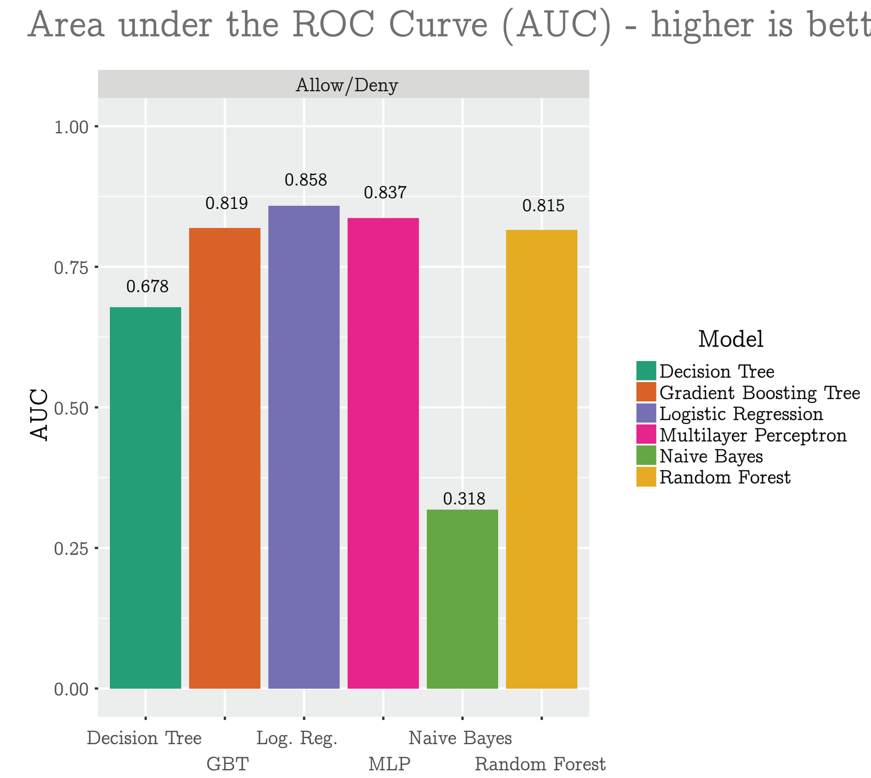


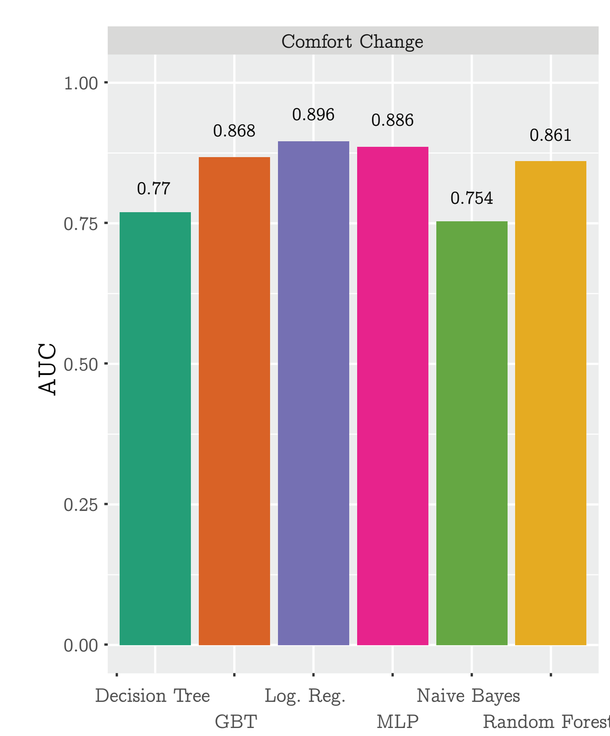
5. Model Comparison (Validation)

Data was ramdomly split 60:30:10, by survey participant ID.
60% training, 30% validation, and 10% test.

Classification Models

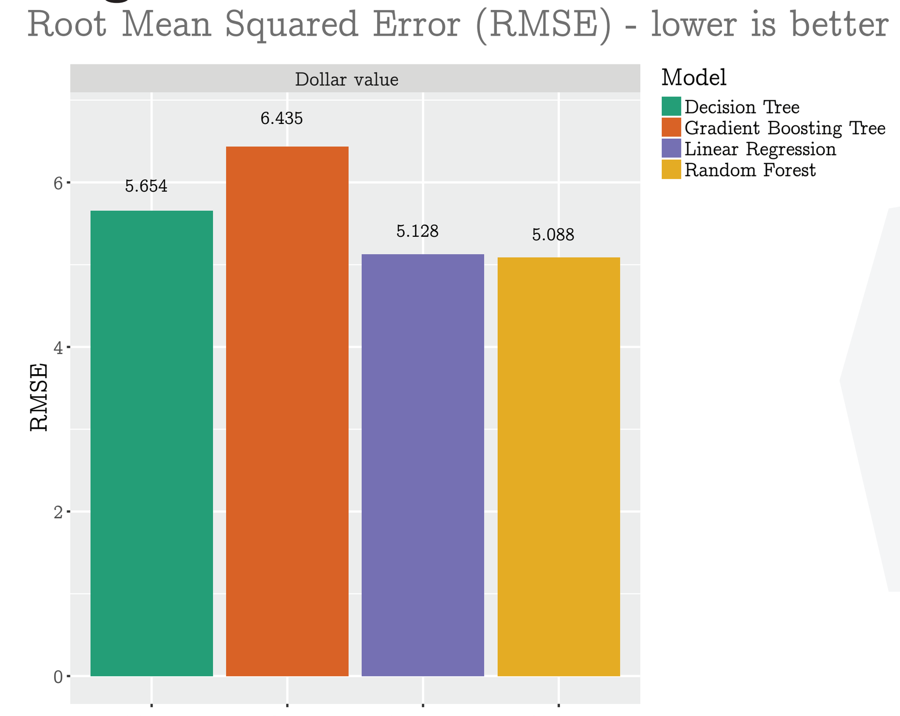
Area under the ROC Curve (AUC) - higher is better





Regression Models

Root Mean Squared Error (RMSE) - lower is better



Logistic regression performed best predicting Allow/Deny and Comfort Change.

Random Forest performed best predicting dollar value.

6. Test Performance and Interpretation

Allow/Deny

AUC=0.870

Predictions on the test set

Purpose		Actual	Predicted
Company Revenue	Allow	4	5
	Deny	8	7
Personalization	Allow	10	16
	Deny	14	8
Home automation	Allow	8	9
	Deny	5	4
Home control	Allow	13	11
	Deny	2	4
Home safety	Allow	12	11
	Deny	8	9
Identity linking	Allow	4	1
	Deny	18	21
Legal actions	Allow	3	2
	Deny	22	23
Price discrimination	Allow	6	4
	Deny	8	10
Targeted ads	Allow	7	6
	Deny	13	14
User tracking	Allow	7	4
	Deny	10	13

Top 5 Coefficients toward Deny

Communications	-4.56
Age of people at home	-4.41
Gender of people at home	-4.10
Destinations	-3.97
Noise levels	-3.67

Top 5 Coefficients toward Allow

Comfort toward manufacturer	1.56
Used for personalization	0.45
Used for home safety	0.31
Comfort if identity involved	0.26
IUIPC Awareness	0.20

Comfort Change

AUC=0.945

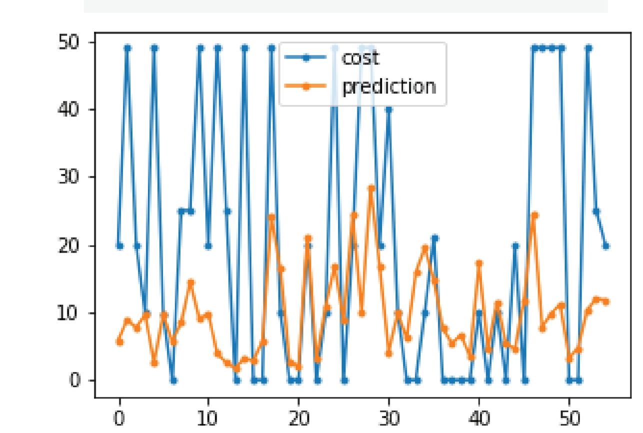
Predictions for the "average user"

Situation	Change	# Scenarios
Ability to Control	More comfortable	11,939
	Less comfortable	4,033
Data secure or not	More comfortable	7,002
	Less comfortable	8,970
Primary vs secondary use	More comfortable	5,791
	Less comfortable	10,181
Aware or not	More comfortable	9,379
	Less comfortable	6,593
Used for safety or not	More comfortable	10,653
	Less comfortable	5,319

Dollar Value

RMSE=4.269

Predictions on the test set



7. Conclusion

By using models such as the ones in this work, manufacturers/developers could derive personalized privacy settings, identify situations in which users would be more or less comfortable, and ultimately learn how much each user is willing to pay for privacy protections.

Future works could explore other techniques that may be effective in predicting the dollar value associated with privacy in the smart home.

Barbosa, N. M., Park, J. S., Yao, Y., & Wang, Y. (2019). "What if?" Predicting Individual Users' Smart Home Privacy Preferences and Their Changes. Proceedings on Privacy Enhancing Technologies, 4, 1-21.

This work was supported in part by the National Science Foundation (NSF) grant number CNS-1464347.