

Comparison of Logit Models to Machine Learning Algorithms for Modeling Individual Daily Activity Patterns

Daniel Fay, Peter Vovsha, Gaurav Vyas
(WSP USA)

Logit vs. Machine Learning Models

Logit Models:

- Convenient model properties
- Easy replication of observed aggregate shares
- Suffer from combinatorial explosion of alternatives
- Mostly linear additive specifications of utilities

Machine Learning Models:

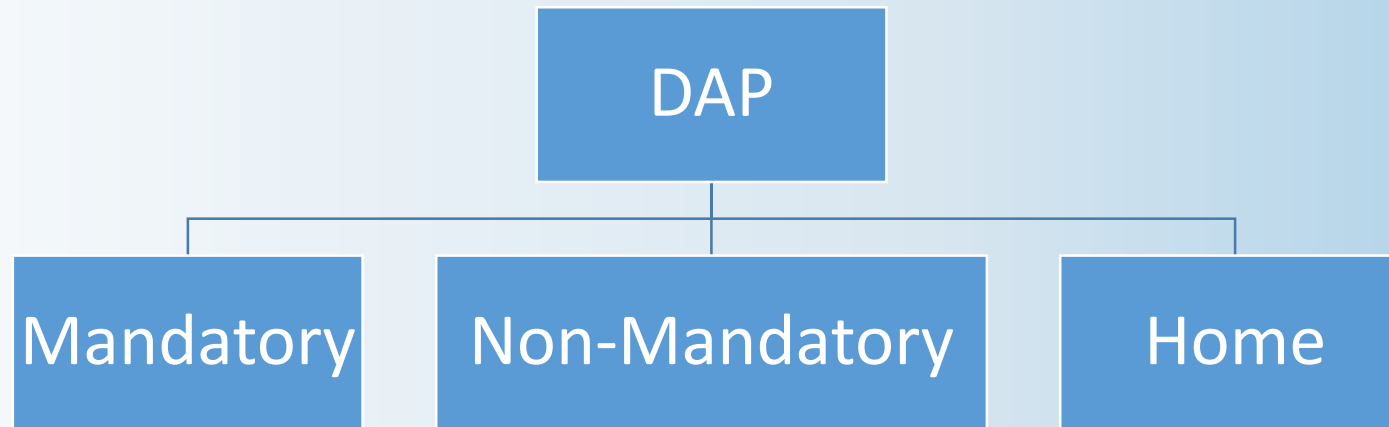
- Capture non-linear affects of variables and their combinations
- Many different ML methods available
- Prioritize individual prediction rather than aggregate shares
- Suffer from systematic over/under predictions

Research Focus

- Individual prediction of daily activity pattern types as part of ABM
- Resolving combinatorial explosion of alternatives
- Applying model constraints to decision trees
- Behavioral insights from combinations of variables provided by decision trees

Individual Daily Activity Pattern Types (DAP)

- 3 categories for each person-day:
 - Mandatory – at least one work, university or school trip
 - Non-mandatory – at least one non-mandatory trip with no mandatory trips
 - Home – no participation in out-of-home activities
- Distinct travel patterns for each type



Modeling Coordinate Daily Activity Patterns

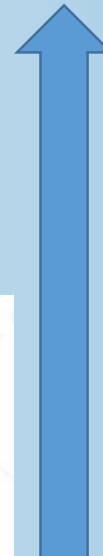
- Important to model DAP type for household members simultaneously
- Trinary choice model applied to household members jointly
- Leads to explosion in number of alternatives

3 Person Family



27 Combinations

7 Person Family



2187 Combinations

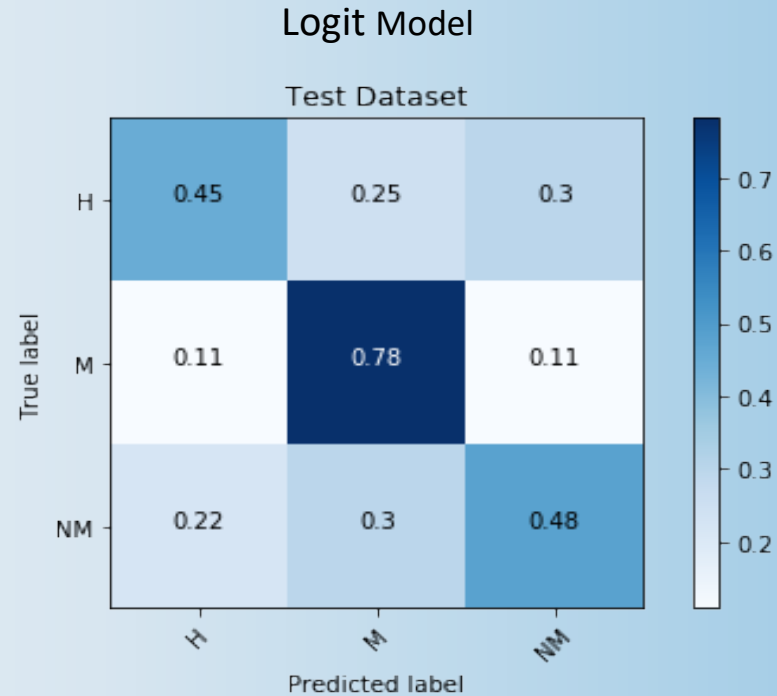
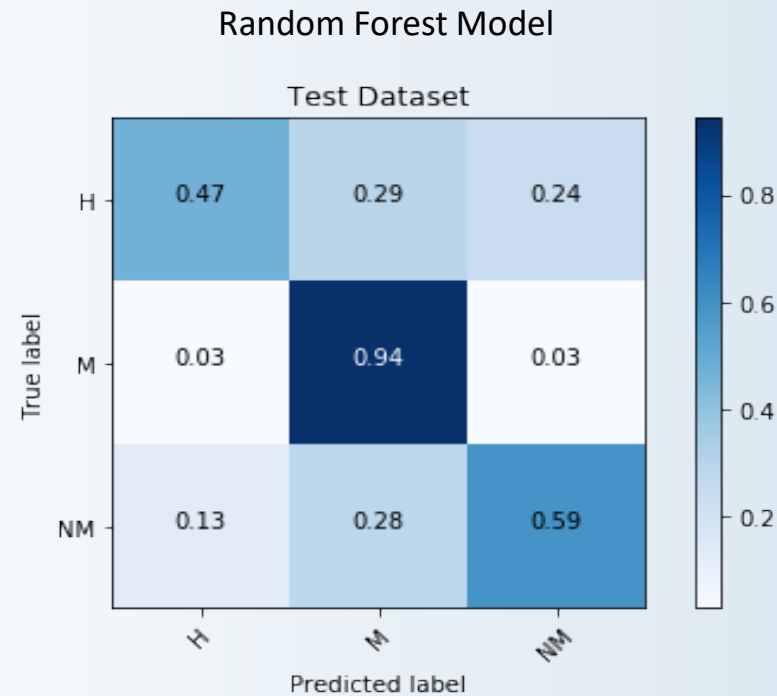
Machine Learning applied to DAP

- Objectives:

- Precision of DAP predicted – individual and aggregate shares
- Find method to resolve combinatorial explosion of set of alternatives
- Identify key variable combinations and the non-linear impacts

Machine Learning applied to DAP

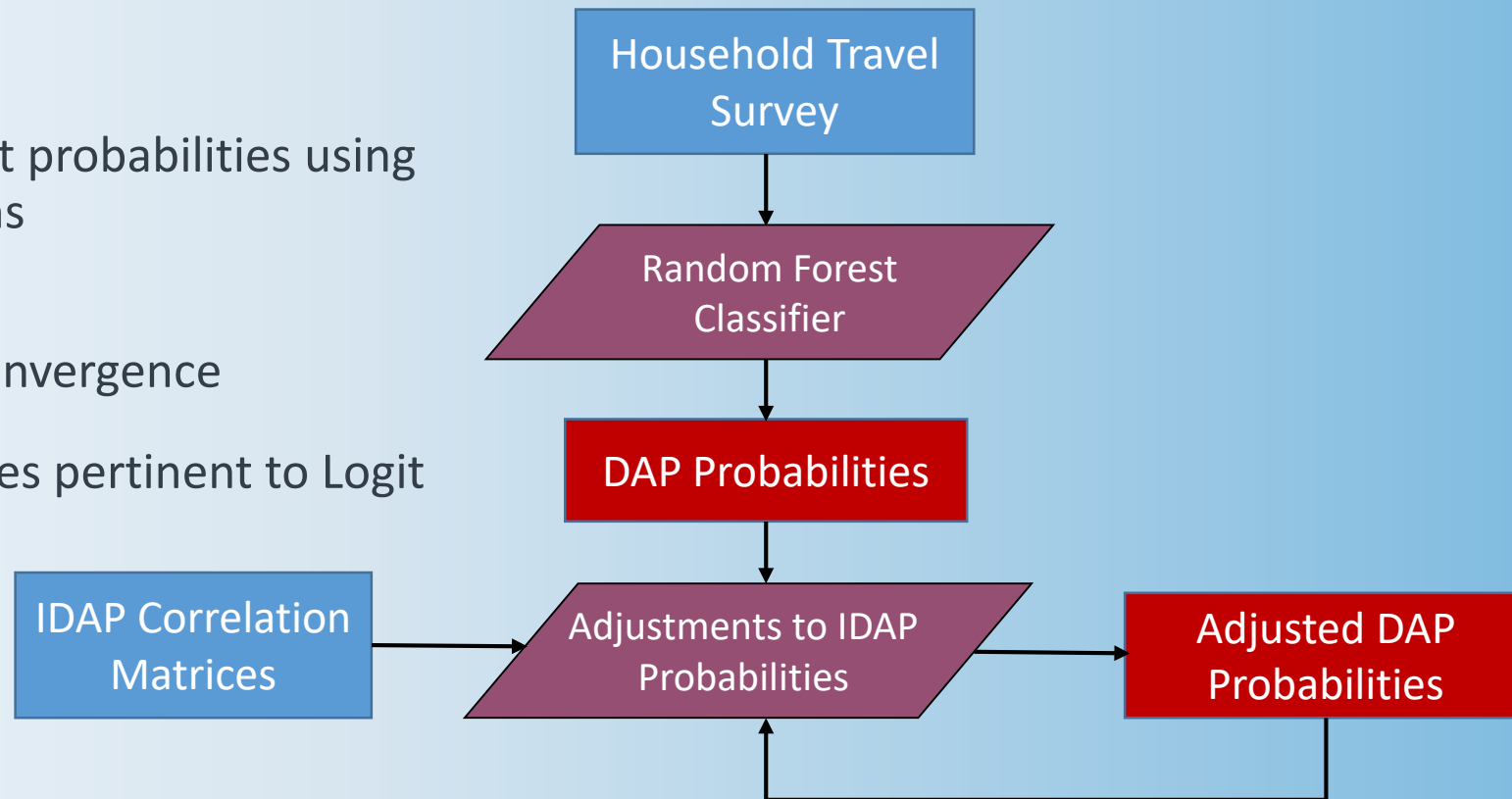
- Individual Accuracy:



Machine Learning applied to DAP

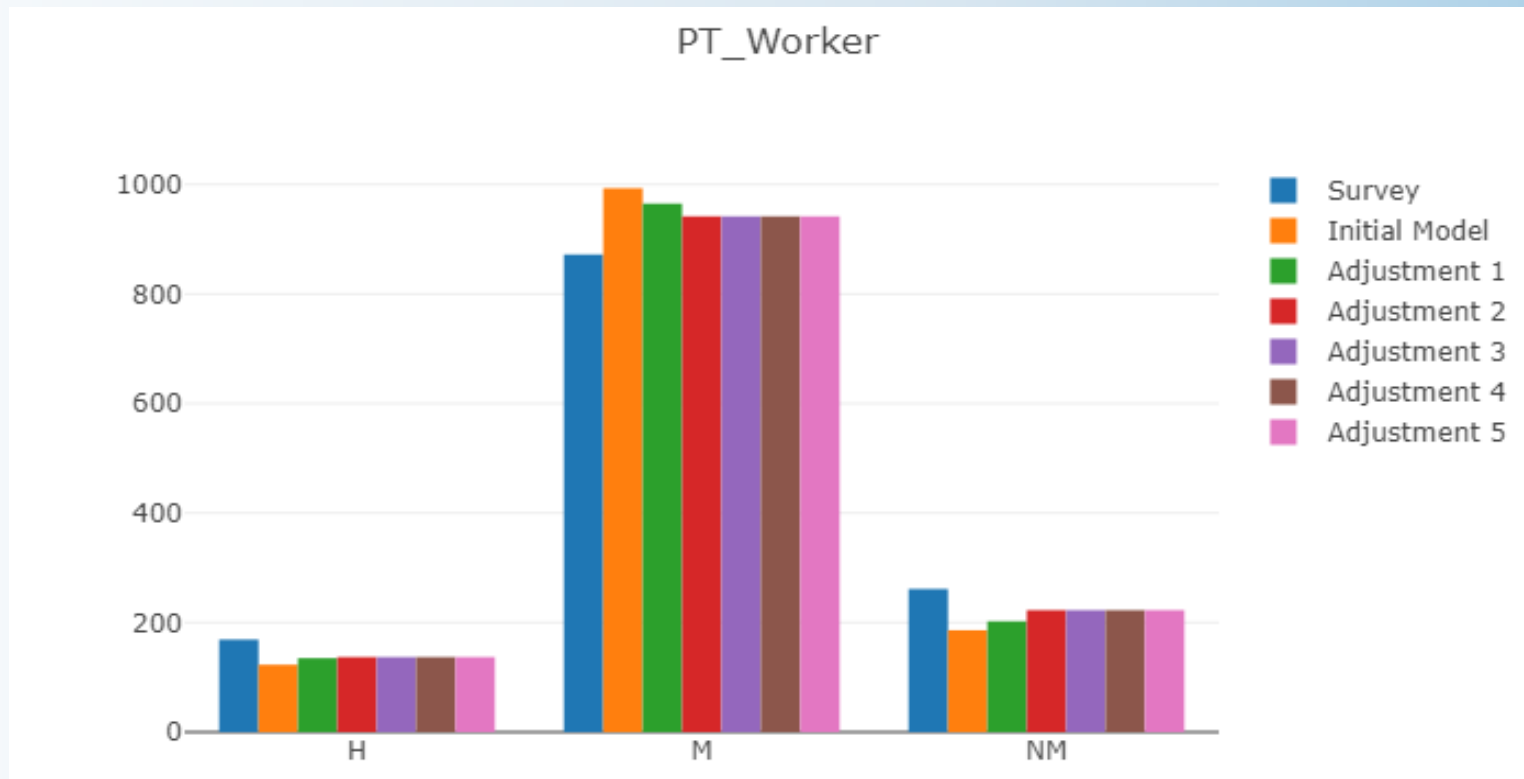
• Resolving Combinatorial Explosion:

- Adjusted initial random forest probabilities using correlations between patterns
 - Pairwise correlations
- Performed iteratively until convergence
- Eliminates explosion of choices pertinent to Logit models

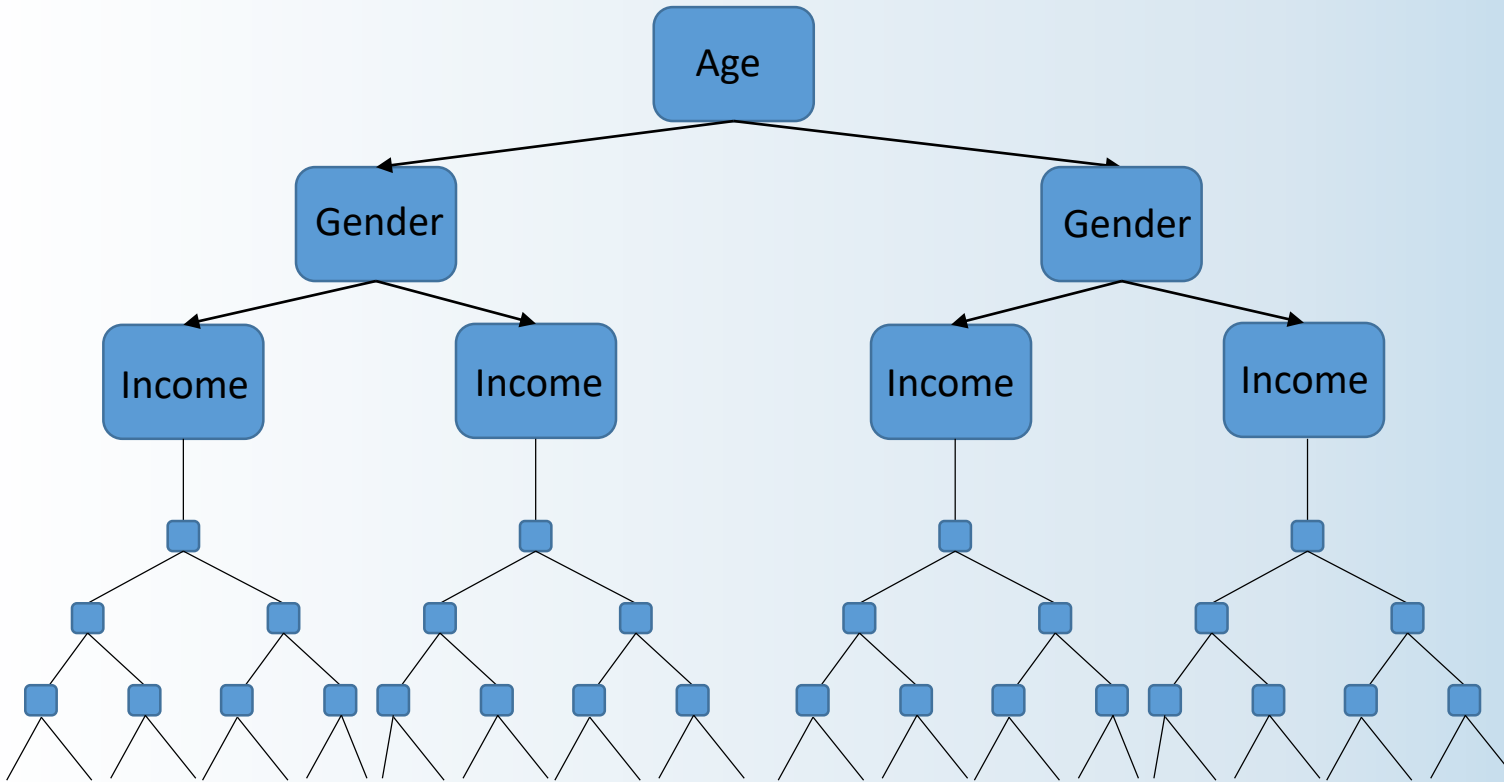


Random Forest Classifier applied to DAP

- Aggregate Accuracy:



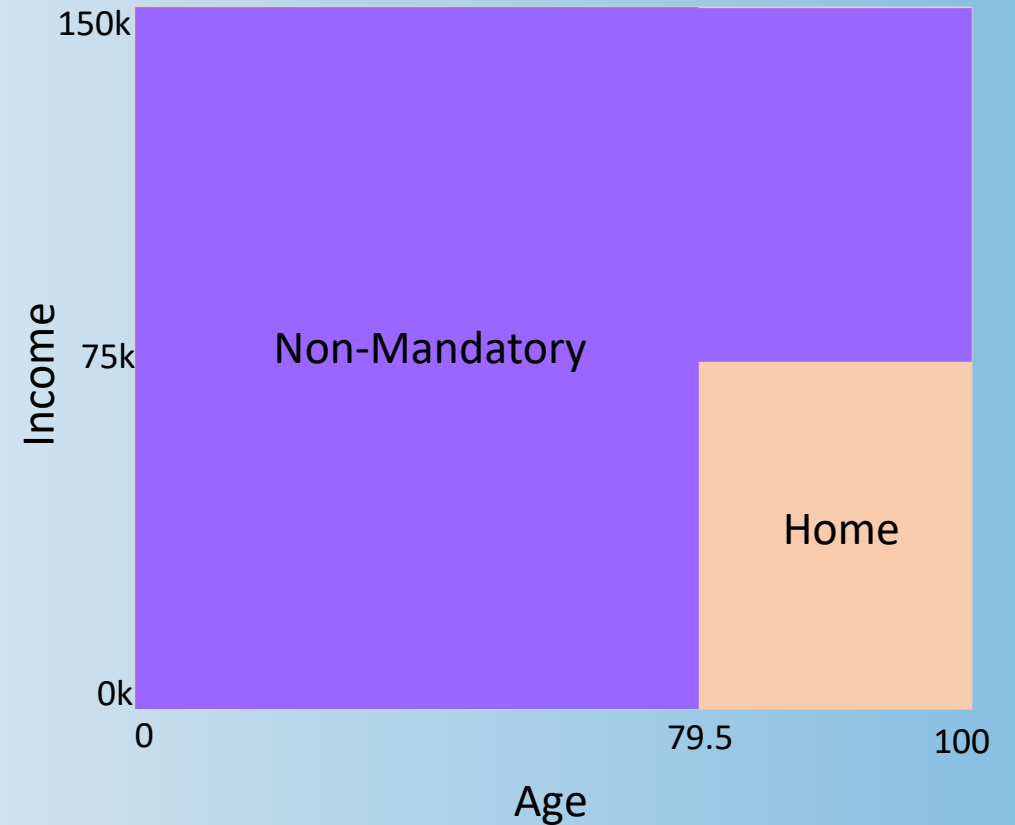
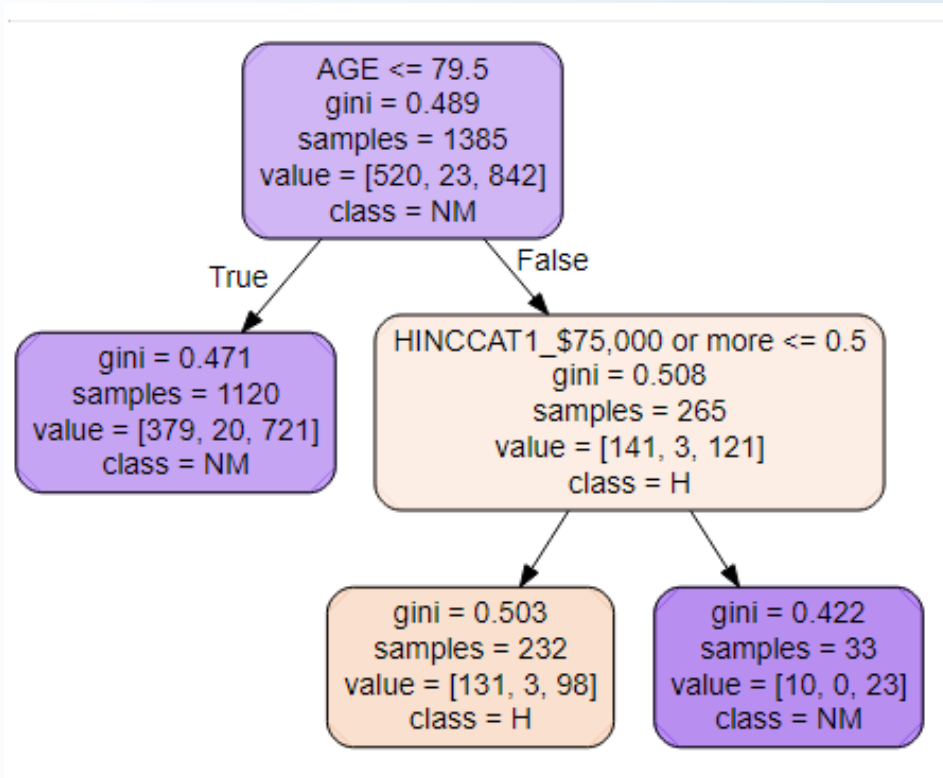
Applying Constraints to Decision Trees to guarantee desired model elasticity



- Constrain first splits of decision tree
- Find optimal split at each leaf node
- Train subsequent branches of the tree

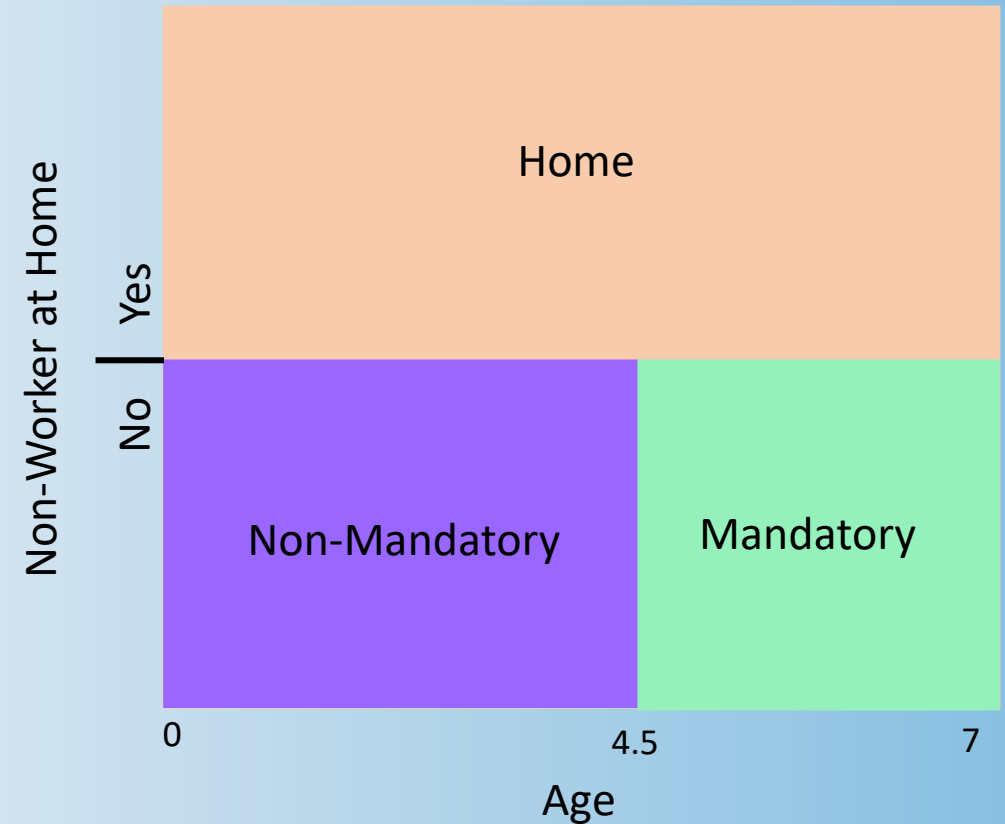
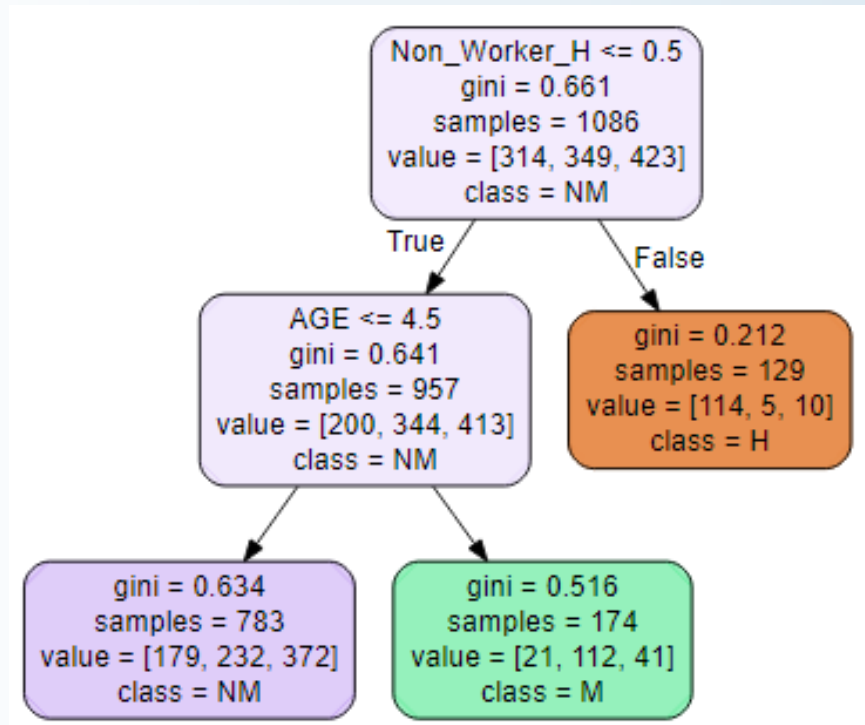
Key Combinations of Variables

Retirees



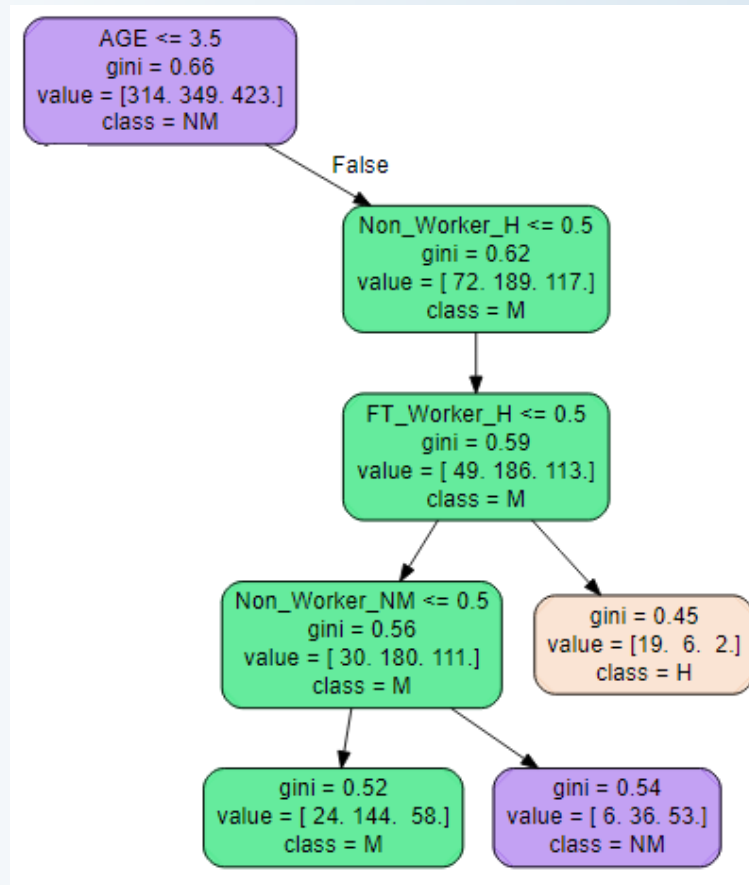
Key Combinations of Variables

Pre-School Children

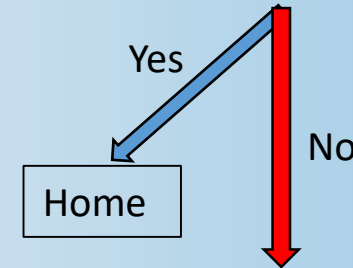


Key Combinations of Variables

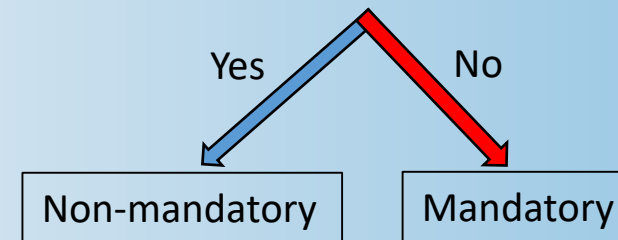
Pre-School Children



- 4 years or older
- No non-worker at home
- Full-time worker at home?

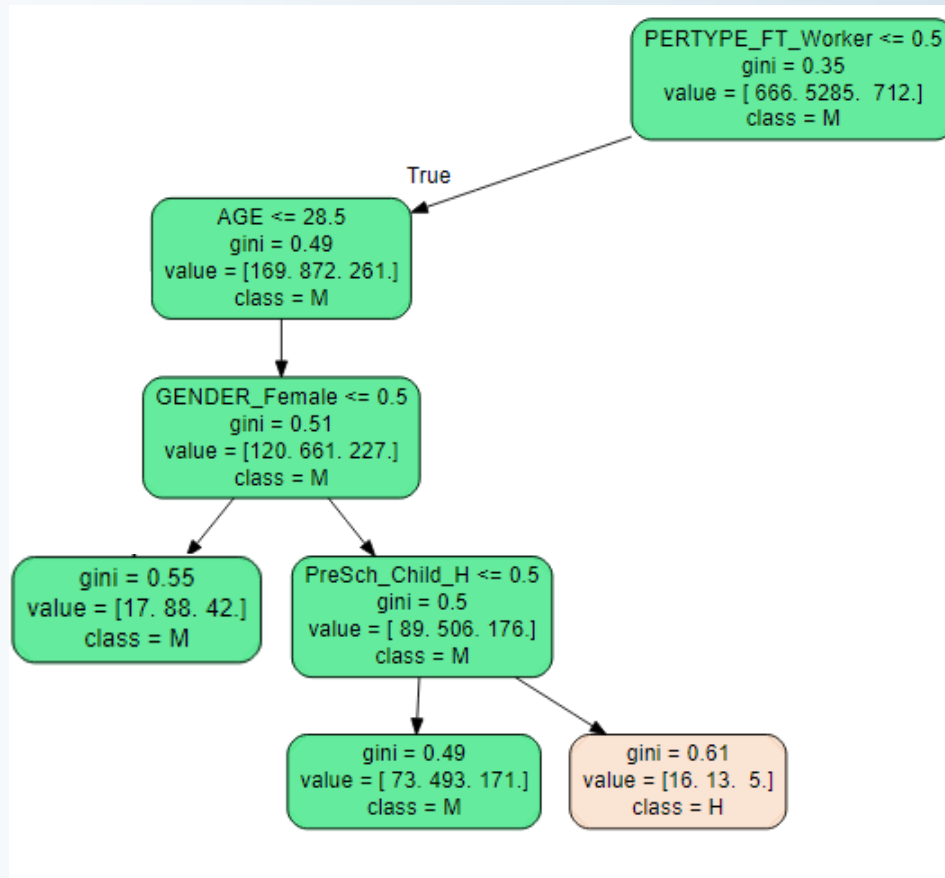


- Non-worker with non-mandatory activity?

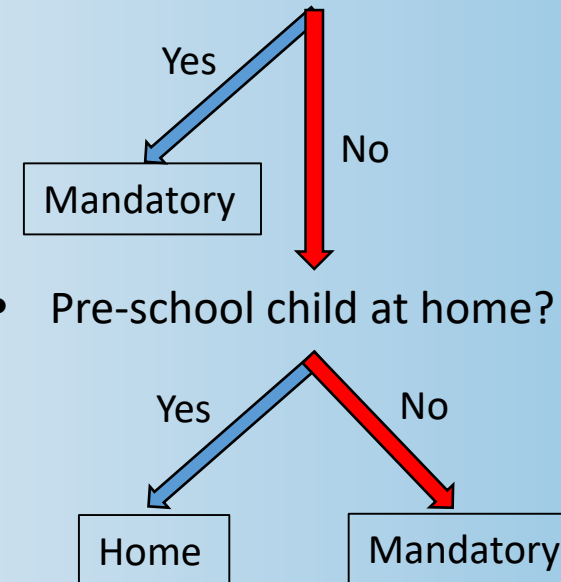


Key Combinations of Variables

Full-Time and Part-Time Workers

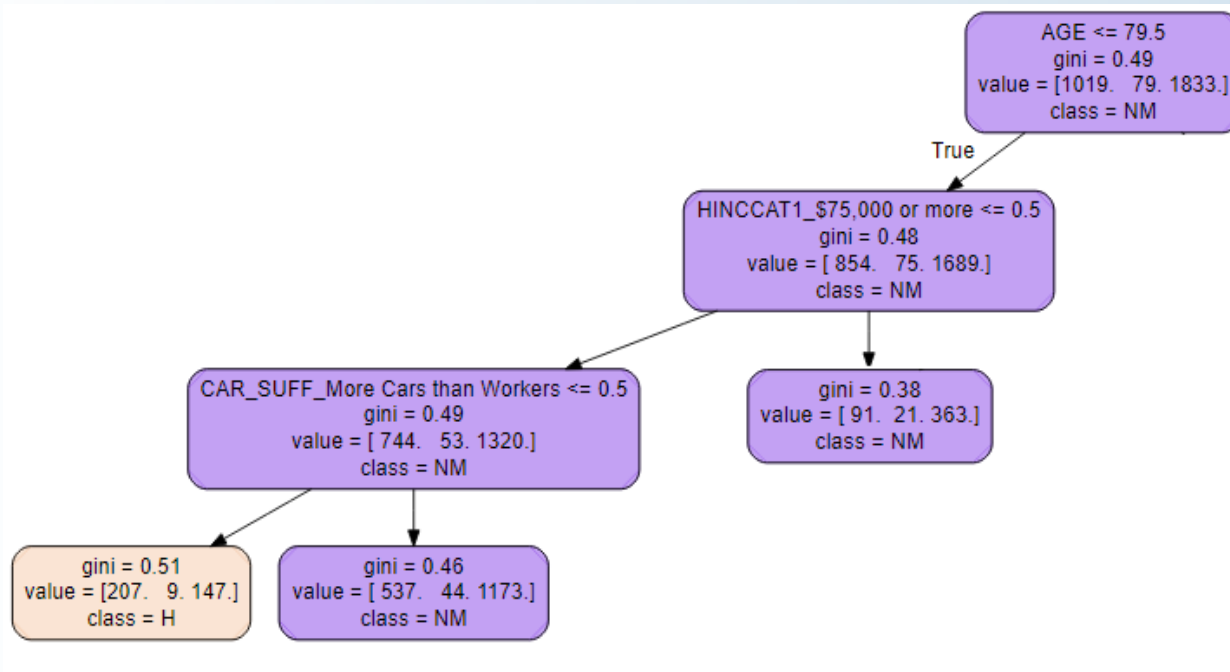


- Part-time worker
- 29 years or older
- Gender Male?

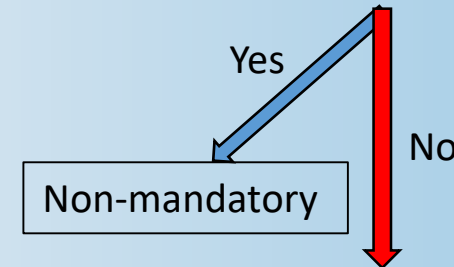


Key Combinations of Variables

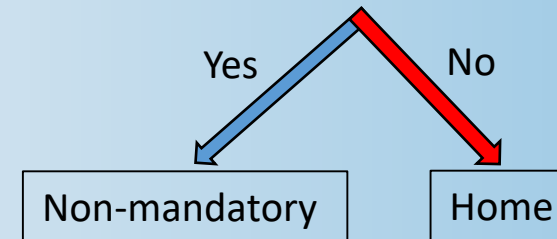
Non-workers and Retirees



- 79 years or younger
- Income 75k or more?



- More cars than workers?



Conclusions

- ML methods represent a viable alternative to traditional logit models for complex multi-dimensional choices. They may improve the individual model fit significantly
- ML may systematically over-predict or under-predict certain choices; in this regard, making ML models easy to calibrate in aggregate sense is an important direction
- ML methods indeed provide some additional insights into travel behavior by revealing certain non-linear combinations of variables that otherwise are difficult to guess and test with traditional logit models
- However some concerns have to be addressed before we can put ML in practice....