19th Behavioral Modelling Summer School
12-15 September, 2020

Application of Machine Learning in Travel Behavior Analysis

Muhammad Awais Shafique, PhD
Assistant Professor
University of Central Punjab, Pakistan

What is Machine Learning?

"Field of study that gives computers the ability to learn without being explicitly programmed."

Arthur Samuel (1959)

"A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E."

Tom Mitchell (1998)

Types of Learning Algorithms

- Supervised learning
  - Teach the computer and then let it use the learning to do the task (labelled data).
- Unsupervised learning
  - Let the computer determine structure and patterns in the data (unlabeled data).
- Semi-supervised learning
  - Let the computer determine patterns and then teach the computer to do the intended task (unlabeled and labelled data).
- Reinforcement learning
  - Let the computer determine the ideal behavior in an environment to maximize its performance.

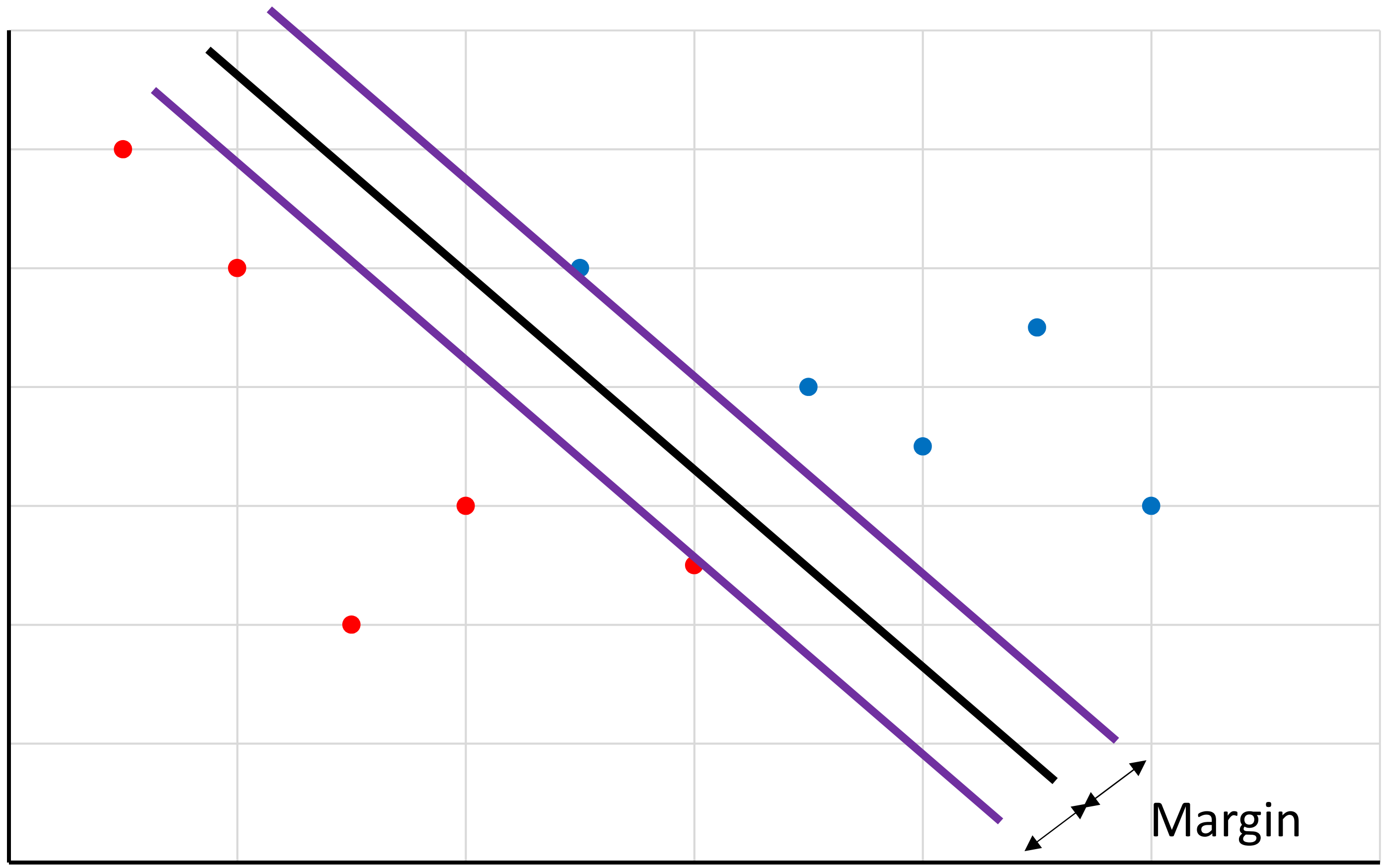
Classification Algorithms

- Support Vector Machines (SVM)
- Decision Tree (DT)
- Random Forest (RF)

Support Vector Machines

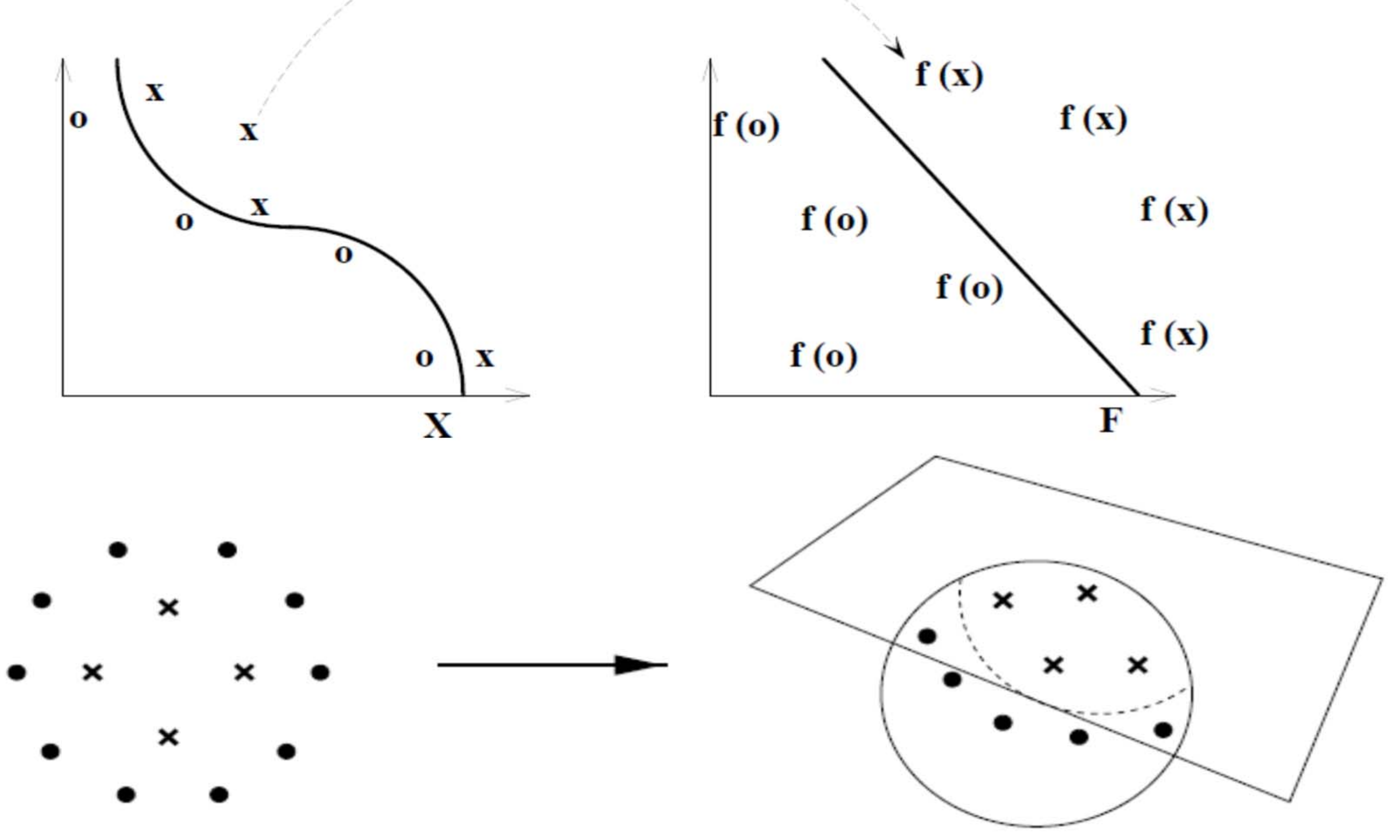
- Among the best “off-the-shelf” supervised learning algorithms.
- Off-the-shelf: A method that can be applied directly to data without requiring a great deal of time-consuming data preprocessing or careful tuning of the learning procedure.
- SVM is a two-class classifier which forms a separating hyperplane.
- When a set of training data containing class labels is supplied to SVM, it outputs an optimal hyperplane which then classifies new examples.

Support Vector Machines



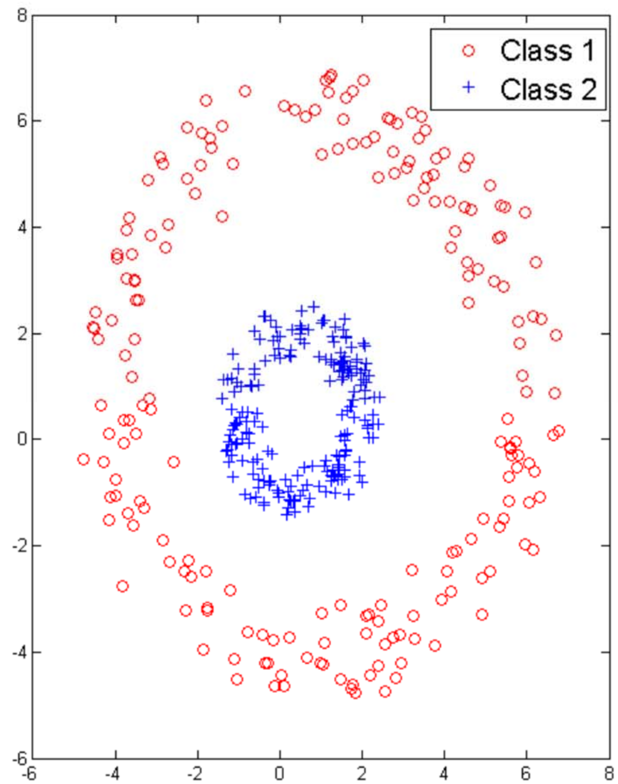
Support Vector Machines

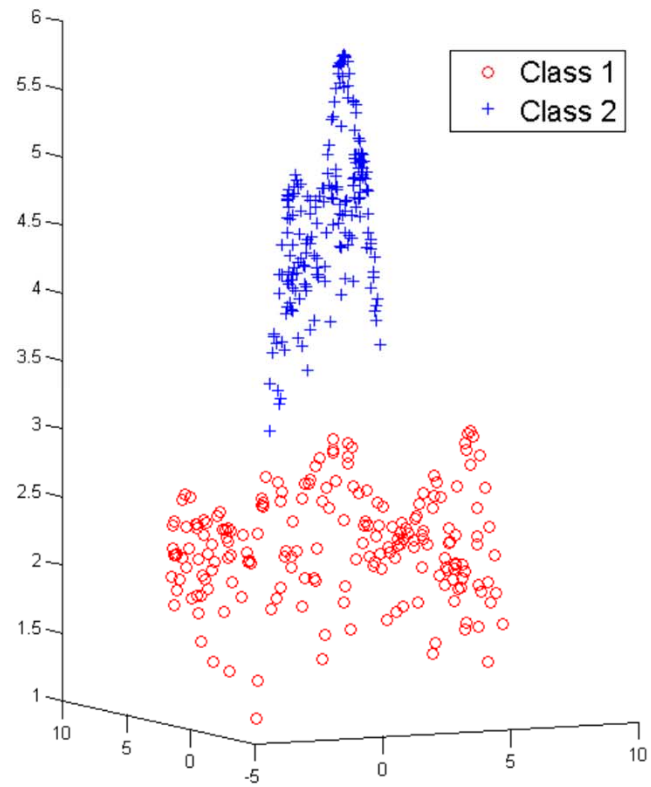
- Map data into a feature space where they are linearly separable.



From Presentations by Nello Cristianini and Jason Weston respectively

Support Vector Machines





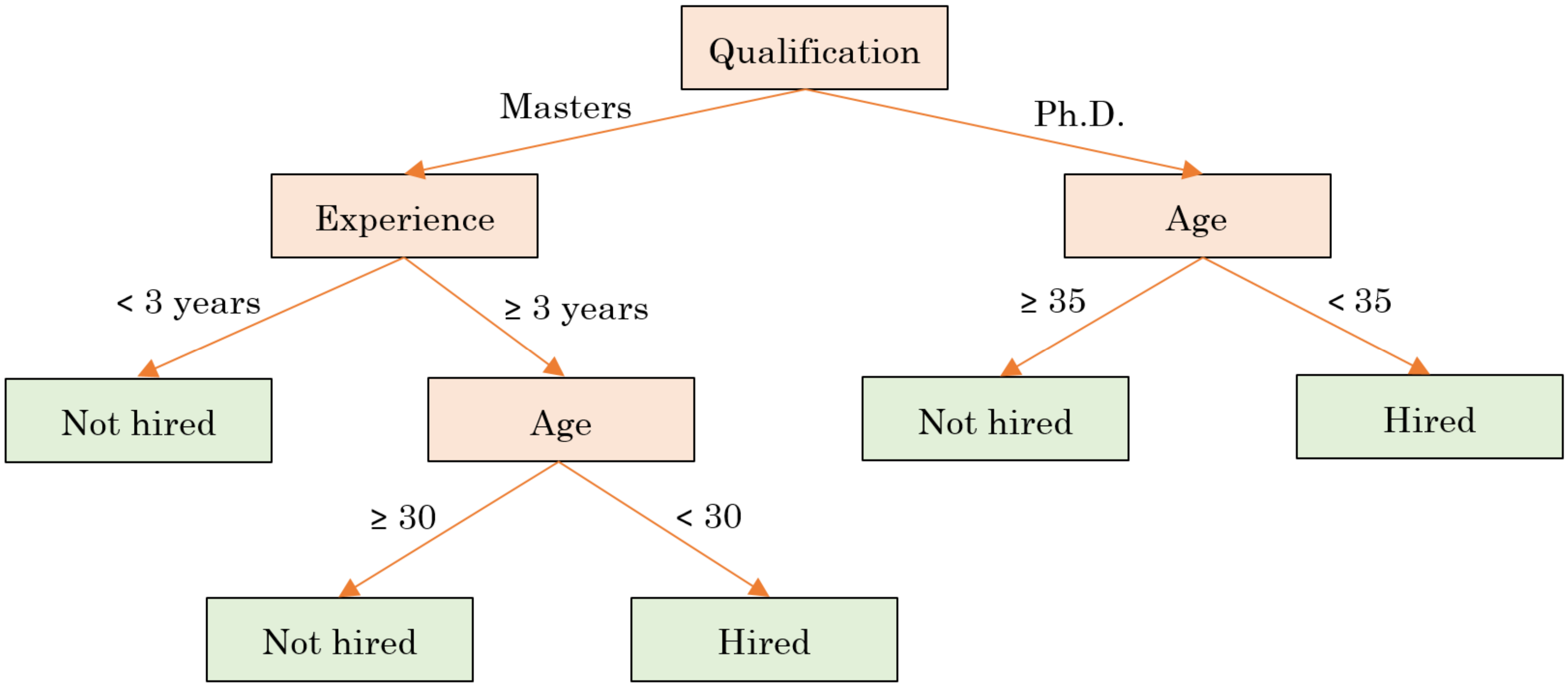
Decision Trees

- Decision trees repeatedly split the dataset in order to arrive at a final outcome.
- The split is made into branch-like segments and these segments progressively form an inverted tree, which originates from the starting node called the root.
- The root is the only node in the tree which does not have an incoming segment.
- The tree terminates at the decision nodes, also known as leaves or terminals.
- All the other nodes present within the tree are called internals or test nodes.

Decision Trees

- The variables or features associated with the data are used to make each split.
- At each node, the variables are tested to determine the most suitable variable to make the split.
- This testing is repeated on reaching the next node and progressively forms a tree.
- Each terminal node corresponds to a target class.

Example


Qualification
Masters
Ph.D.
Experience
Age
< 3 years
≥ 3 years
≥ 35
< 35
Not hired
Age
Not hired
Hired
≥ 30
< 30
Not hired
Hired

Random Forest

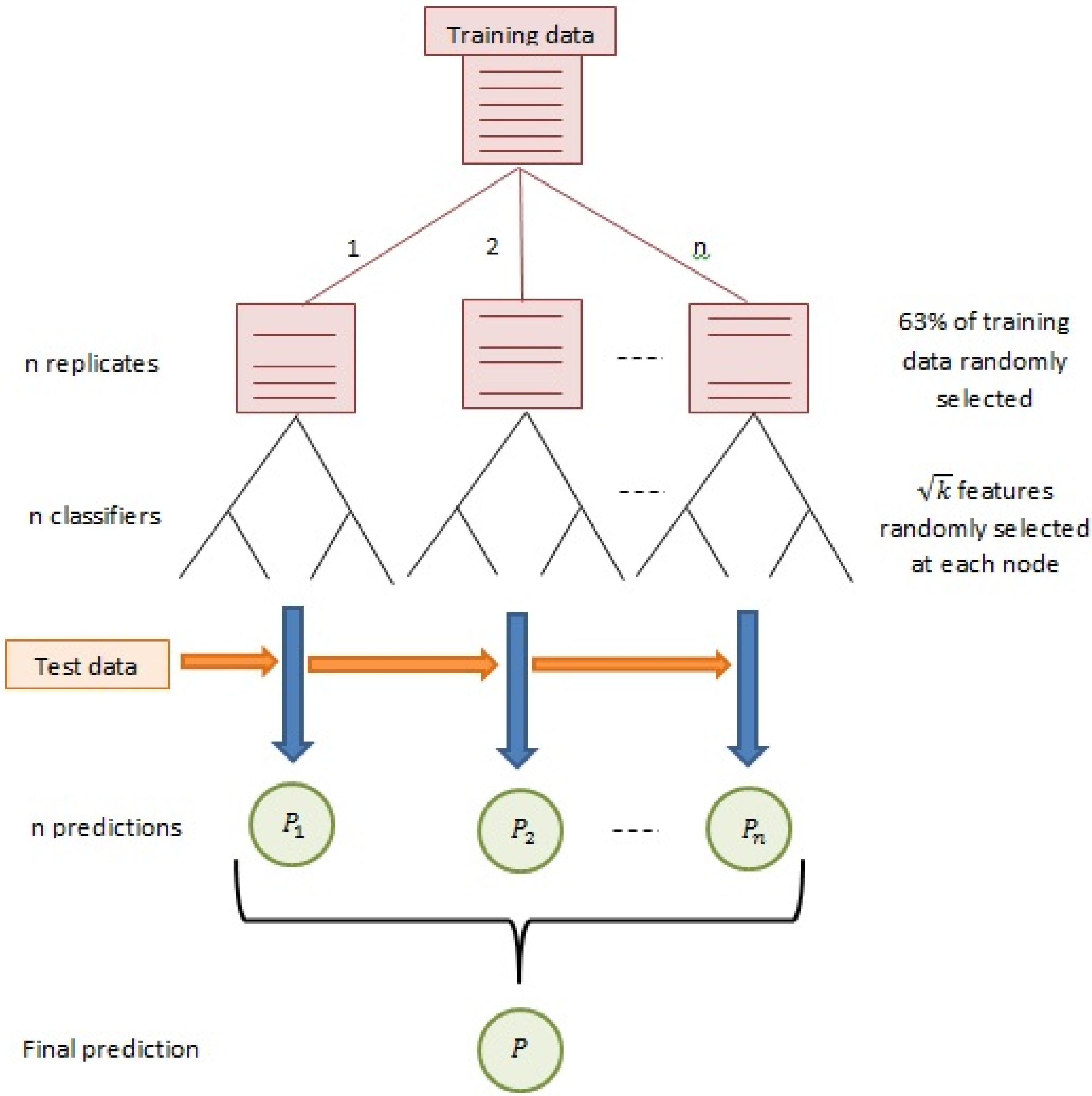
- A bootstrap sample of the dataset is used to independently grow individual trees, and the majority vote is taken to conclude the final prediction.
- In addition to using a randomly selected bootstrap sample of the data for growing each tree in the forest, randomness was introduced in the splitting of nodes.
- In standard trees, each node in the classification or regression tree is split using the best split among all variables.
- In random forests, each node is split using the best among a subset of the variables randomly selected at that node.

Random Forest

- Suppose n number of trees are grown. Each tree is generated by randomly selecting nearly 63% of the given training data.
- The sample data is therefore different for each tree. The remaining 37% data, known as out of bag (OOB) data, is used to estimate the error rate.
- The trees are fully grown without any requirement of pruning, which is one of the advantages of random forest.

Random Forest

- At each node a subset of variables or features is selected and the most suitable feature among them is used for the split.
- The size of subset is a variable which is generally taken as $\sqrt{k}$ where k is the total number of features.
- Once the forest is grown by using the labelled training dataset, the test data is introduced for the prediction.
- The individual predictions by the trees are aggregated to conclude the final prediction result (i.e. majority vote for classification and average for regression).


Training data
1
2
n
n replicates

63% of training data randomly selected
n classifiers

$\sqrt{k}$ features randomly selected at each node
Test data
n predictions
P_1
P_2

P_n
Final prediction
P

Discussion Paper

- Zhao, X., Yan, X., Yu, A., & Van Hentenryck, P. (2020). **Prediction and behavioral analysis of travel mode choice: A comparison of machine learning and logit models.** *Travel behaviour and society*, *20*, 22-35.

Introduction

- For travel-behavior analysis, we mostly use discrete choice modelling.
- Particularly the use of models belonging to logit family is quite popular such as Multinomial Logit Model, Nested Logit Model and Mixed Logit Model.
- With tremendous increase in popularity of machine learning, it is finding its way into travel-behavior research as well.

Model Development

- Logit models are based on calculating utilities (deterministic and error term) for available alternatives, and selecting the alternative which gives maximum utility.
- They have a defined layered structure.
- In contrast, machine learning methods learn from the data to form a target function that maps the input variables to the target (alternative selected).
- The structure is quite flexible, and may follow layered structure, tree-based structure, rule-based structure etc.

Model Evaluation

- Goal: Reduce overall prediction error (bias + variance + irreducible error)

- Bias: Error due to incorrect assumptions
- Variance: Error due to model sensitivity
- Irreducible error: Error due to data noise

- Bias-variance tradeoff → underfitting vs overfitting

Data Collection

- Stated-preference survey conduced in University of Michigan.
- 8,141 observations collected from 1,163 individuals.
- Respondents were asked to estimate trip attributes.
- Questions were then asked about a new public transit system.
- Variables included trip attributes, socio-demographic aspects, residential preferences, and selected mode choices.

Data Collection

Travel time of driving (min)
Travel time of walking (min)
Travel time of biking (min)
Travel time of using PT (min)
Parking cost ($)
Wait time for PT (min)
Number of transfers
Number of additional pickups
Income level
Importance of bike- and walk-ability
Importance of PT access
Car per capita
Female or male

Models Examined

- Multinomial Logit Model
- Mixed Logit Model
- Naïve Bayes (NB)
- Classification and Regression Trees (CART)
- Random Forest (RF)
- Boosting Trees (BOOST)
- Bagging Trees (BAG)
- Support Vector Machines (SVM)
- Neural Networks (NN)

Result Comparison

Mean out-of-sample accuracy of machine-learning and logit models (individual level).

Model	All		Car		Walk		Bike		PT	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
MNL	0.647	0.016	0.440	0.044	0.859	0.018	0.414	0.033	0.698	0.029
Mixed logit	0.631	0.008	0.513	0.031	0.797	0.014	0.413	0.038	0.673	0.027
NB	0.584	0.018	0.558	0.035	0.864	0.013	0.372	0.041	0.490	0.042
CART	0.593	0.014	0.428	0.032	0.795	0.022	0.329	0.038	0.653	0.026
BOOST	0.850	0.007	0.790	0.035	0.913	0.012	0.848	0.023	0.825	0.028
BAG	0.854	0.013	0.791	0.017	0.926	0.016	0.861	0.028	0.818	0.029
RF	0.856	0.012	0.797	0.022	0.928	0.016	0.859	0.021	0.820	0.027
SVM	0.772	0.012	0.701	0.027	0.878	0.026	0.681	0.033	0.770	0.026
NN	0.646	0.016	0.434	0.045	0.853	0.025	0.451	0.051	0.679	0.024

Result Comparison

Mean L1-Norm error for mode share prediction.

Model	Mean	SD
MNL	0.0399	0.0207
Mixed logit	0.0593	0.0268
NB	0.2771	0.0363
CART	0.0463	0.0280
BOOST	0.0291	0.0151
BAG	0.0253	0.0130
RF	0.0248	0.0128
SVM	0.0362	0.0218
NN	0.0493	0.0196

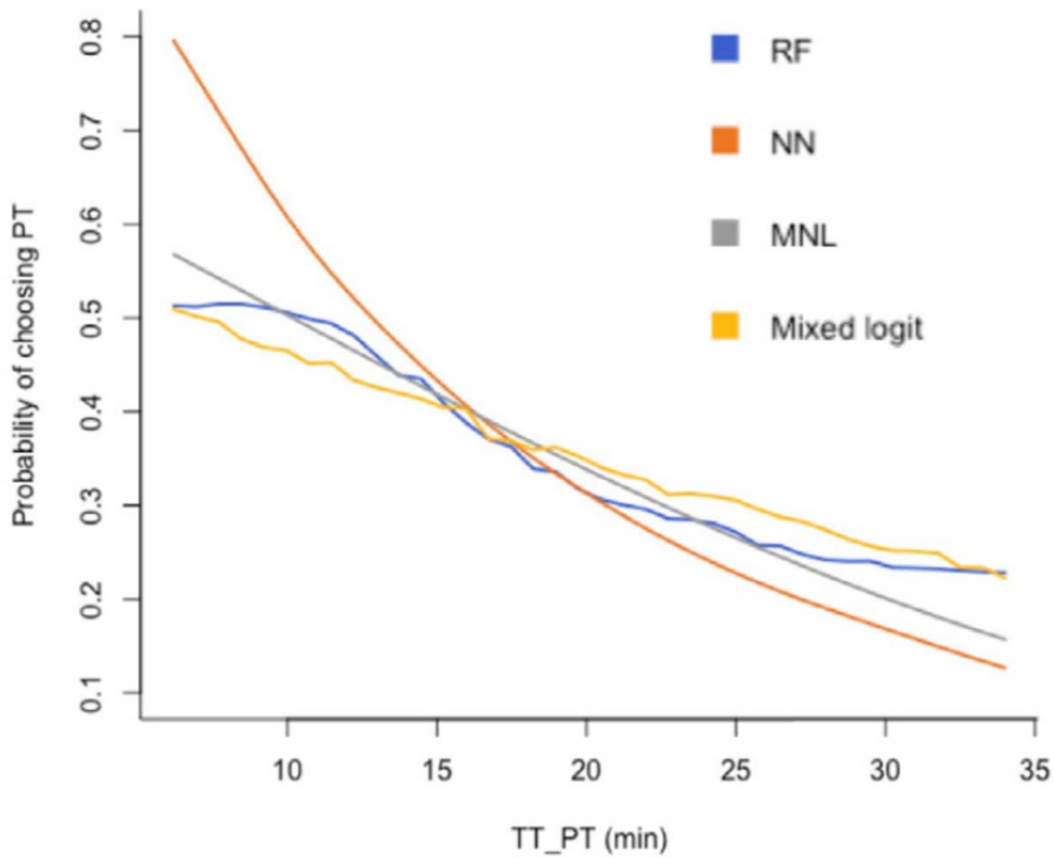
Model Interpretation

Ranking of variable importance for RF, NN, MNL, and mixed logit.

Variable	RF	NN	MNL	Mixed logit
TT_Walk	1	16	2	2
TT_Drive	2	14	4	5
TT_Bike	3	13	1	1
TT_PT	4	11	3	3
Current_Mode_Bike	5	2	5	10
PT_Access	6	8	13	16
Bike_Walkability	7	6	16	13
Income	8	10	14	15
CarPerCap	9	7	11	14
Current_Mode_Walk	10	1	6	12
Rideshare	11	9	9	9
Transfer	12	5	8	8
Wait_Time	13	15	10	11
Female	14	3	15	17
Parking_Cost	15	12	12	4
Current_Mode_Car	16	4	7	6
Current_Mode_PT	/	/	17	7

Model Interpretation

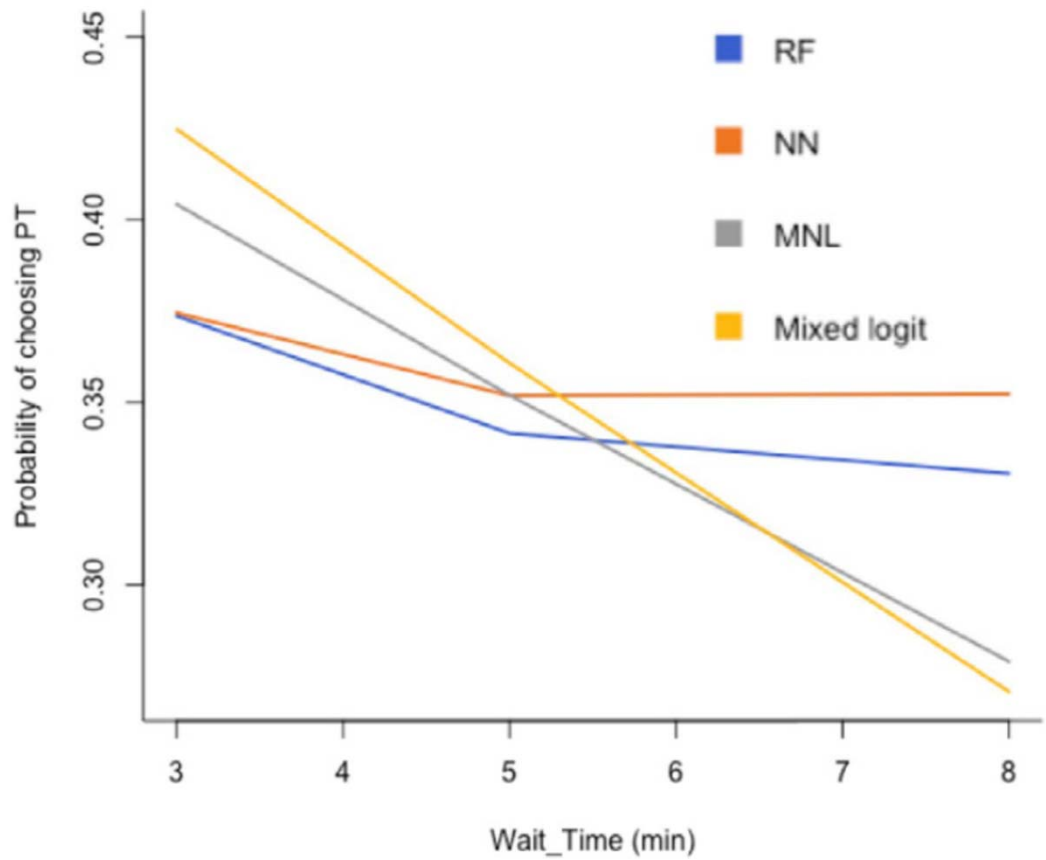
RF has relative flat tails before 10 min and after 25 min, showing people tend to become insensitive to very short or very long transit times



(a) Partial dependence on TT_PT

Model Interpretation

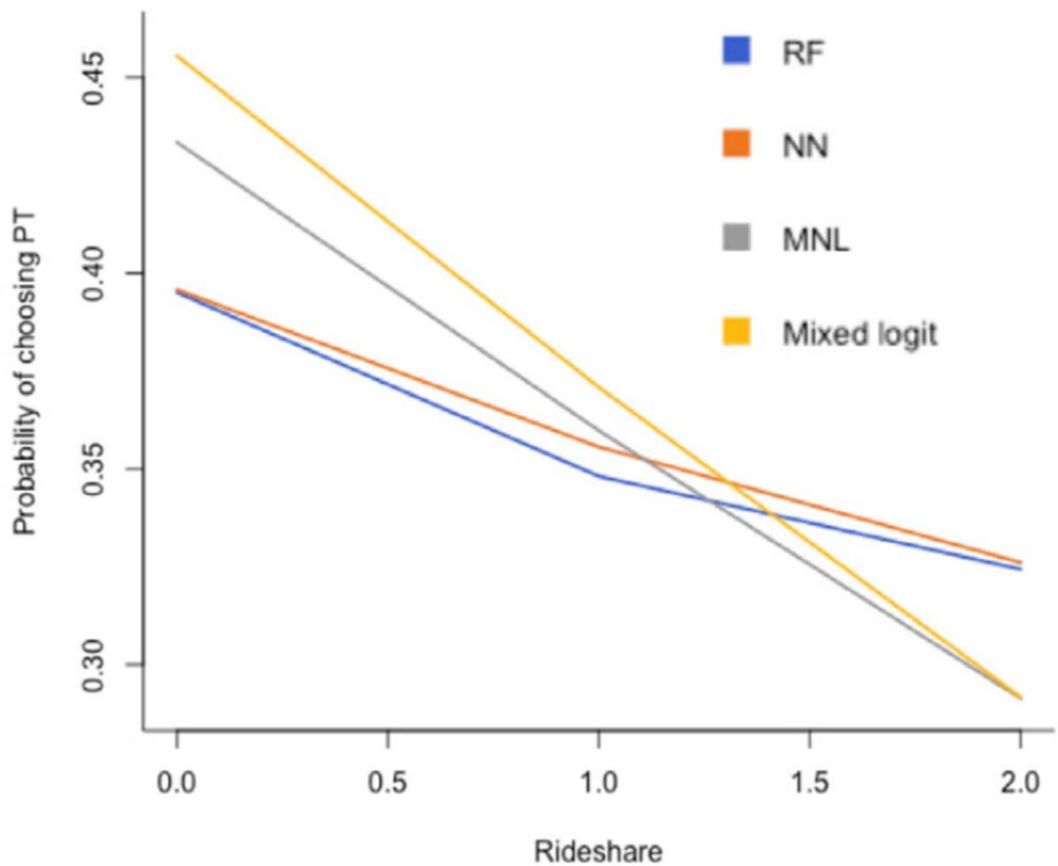
Travelers are more sensitive to wait times less than 5 min



(b) Partial dependence on Wait_Time

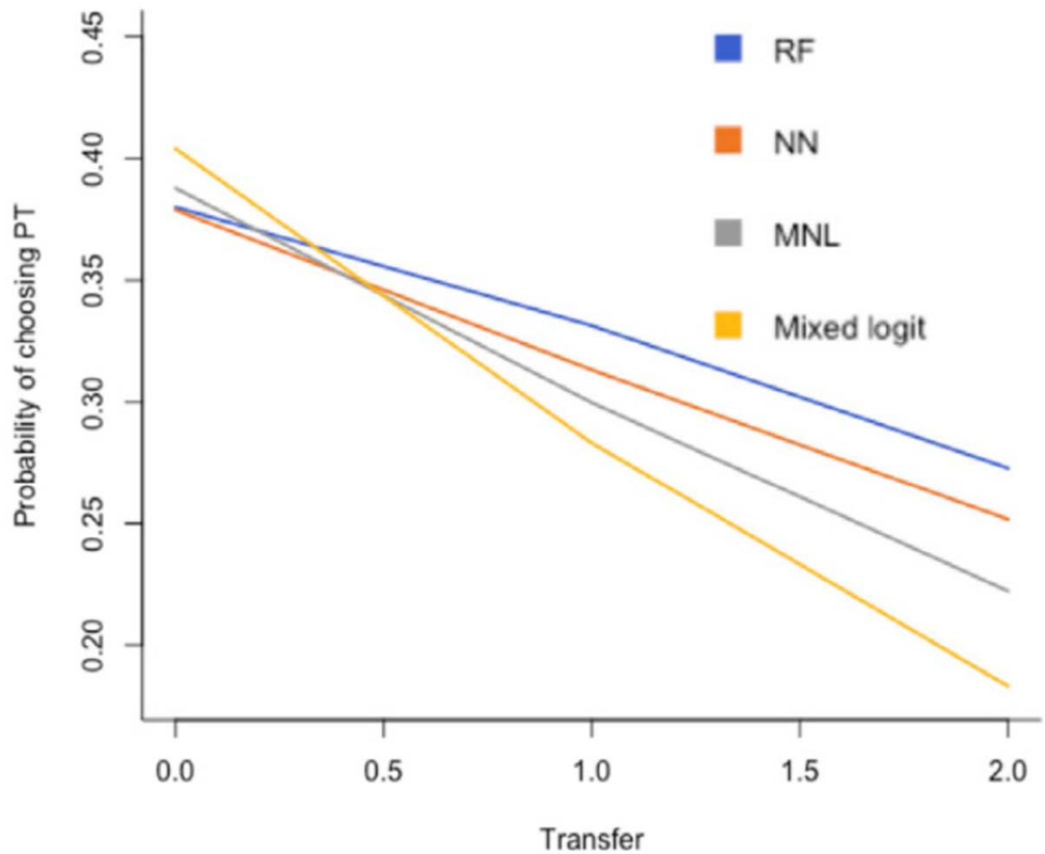
Model Interpretation

The choice probability of PT decreases more significantly from 0 to 1 rideshare compared to from 1 to 2 rideshares.



(c) Partial dependence on Rideshare

Model Interpretation



(d) Partial dependence on Transfer

Model Interpretation

- Consequently, different coefficients for a variable in different data intervals were specified for the logit models (MNL and mixed logit).
- The model fit improved and the coefficient estimates largely agreed with the nonlinearies revealed by the RF model.
- Thus, ML Models can automatically learn the nonlinearities.
- Can assist in improving Logit Models.

Model Interpretation

Marginal effects and arc elasticities.

Variable		Δ	MNL	Mixed logit	NN	RF	
						Standard Approach	Modified Approach
Wait_Time	Marginal effect	1 or 2 min	− 2.93%	− 2.96%	− 0.58%	− 0.01%	− 1.16% [a]
Transfer	Marginal effect	1 unit	− 10.69%	− 11.66%	− 6.27%	− 4.60%	− 5.10%
Rideshare	Marginal effect	1 unit	− 8.13%	− 7.74%	− 2.54%	− 2.08%	− 3.41%
TT_PT	Marginal effect	1 min	− 1.94%	− 0.87%	− 2.45%	− 1.63%	− 1.63%
	Arc elasticity	10%	− 0.89	− 0.49	− 1.28	− 1.07	− 1.08

Note: 2 min is used for the modified case of RF, while 1 min is used for other cases.

Model Interpretation

Penalty of	MNL	Mixed Logit	NN	RF
Transfer	5.5 min	13.4 min	2.6 min	3.1 min
Rideshare stop	4.2 min	8.9 min	1.0 min	2.1 min
Wait time	1.5 min	3.4 min	0.2 min	0.7 min

According to literature

- Penalty effects of a transfer > 5 min of in-vehicle travel time
- Value of wait time is slightly larger than that of in-vehicle travel time

Model Interpretation

- The behavioral outputs of the logit models appear to be more reasonable than those of RF.

- Another possibility: the results of RF are closer to ground truth.
- In-vehicle time vs total time
- Proposed mobility-on-demand system is described as an app-based system that provides accurate real-time information. This may lead to smaller penalties.

Conclusion

- Random Forest significantly outperforms the logit models in prediction both at the individual level and the aggregate level.
- Mixed logit model underperforms the MNL.
- Machine-learning and logit models largely agree on variable importance and the direction of influence that each independent variable has on the choice outcome.
- Partial dependence plots show that machine-learning models can readily capture nonlinear associations.
- RF’s behavioral results are somewhat inconsistent with the existing literature.

awais.shafique@ucp.edu.pk

facebook.com/**m.awaisshafique**

Jp.linkedin.com/in/awaisshafique/

www.researchgate.net/profile/Muhammad_Shafique19