Neurosymbolic Approach for Travel Demand Prediction: Integrating Decision Tree Rules into Neural Networks

Kamal Acharya*, Mehul Lad*, Liang Sun#, Houbing Song*

*University of Maryland Baltimore County, #Baylor University

IWCMC

April 30, 2026

Contents

- Accurate travel demand prediction is essential for transportation planning, congestion reduction, and infrastructure investment.
- Traditional models (e.g., statistical, time-series) struggle to capture nonlinear and multidimensional interactions in travel data.
- Neural Networks (NNs) offer high predictive power but lack interpretability, crucial for transportation decision-making.

- We propose a Neurosymbolic AI approach integrating Decision Tree (DT) rules with NNs.
- DTs extract interpretable *if-then* rules, while NNs model complex travel demand patterns.
- Encoding DT rules as additional features in NNs bridges symbolic reasoning and deep learning.
- Our approach improves accuracy (MAE, R^2, CPC) while enhancing interpretability.

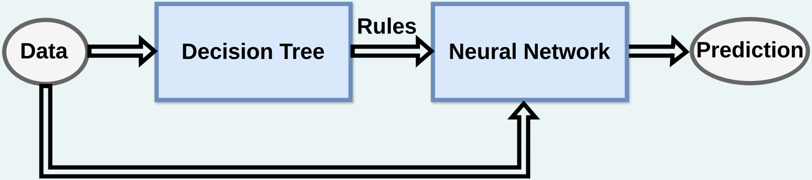


Figure: Neurosymbolic Approach Used in the Research

Study	Methodology	Key Findings
[1]	Logistic regression	Used for travel demand estimation but limited in capturing complex interactions.
[2]	Neural Networks	Effectively captures nonlinear relationships in mobility data.
[3]	Tree-based models + Interpretation techniques	Boosting trees outperform other models, with feature importance varying based on hyperparameters.
[4]	Extreme Gradient Boosting (XGB)	Identifies key factors influencing travel mode choice using variable importance and interaction analysis.
[5]	SHAP analysis + Variable importance	Explains travel mode choice by identifying influential factors such as trip distance, income, and car ownership.

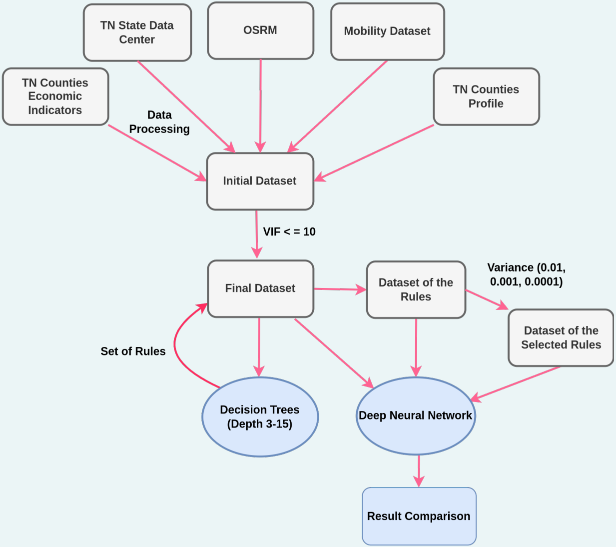


Figure: Research Process

Table: County-Level Features in Tennessee

Feature Category (Count)	Description
Land Use Counts (2)	NaturalAreaCounts (forests, agriculture), PublicAreaCounts (residential, commercial, industrial)
Points of Interests (1)	Counts of educational institutions, commercial areas, healthcare facilities, entertainment centers
Roads (2)	MajorRoads (highways, expressways), OtherRoads (local, tertiary roads)
Transportation (1)	Number of airports, railway stations, bus stations
Economic Features (3)	Includes Unemployment Rate, Employed Population, and Sales Tax Revenue
Ranking (1)	Overall ranking of counties based on health, economic well-being, and education
Population (1)	Total population count for each county

Table: 8 Rules Extracted from Decision Tree of Depth 3

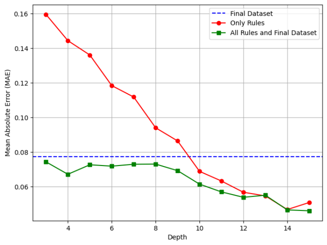
Rules
distance_miles $\leq$ 46.08 AND POIs_Destination $>$ 323.0 AND POIs_Origin $>$ 307.0
distance_miles $\leq$ 46.08 AND POIs_Destination $>$ 323.0 AND POIs_Origin $\leq$ 307.0
distance_miles $\leq$ 46.08 AND 323.0 $\geq$ POIs_Destination $>$ 243.0
46.08 $<$ distance_miles $\leq$ 58.77 AND NaturalAreaCounts_Destination $>$ 935.5
distance_miles $\leq$ 46.08 AND POIs_Destination $\leq$ 243.0
distance_miles $>$ 58.77 AND POIs_Destination $>$ 773.5
46.08 $<$ distance_miles $\leq$ 58.77 AND NaturalAreaCounts_Destination $\leq$ 935.5
distance_miles $>$ 58.77 AND POIs_Destination $\leq$ 773.5

Table: Rules and Variance-Based Rules at Different Tree Depths

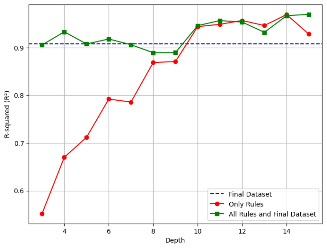
Depth	All Rules	Rules Selected by Variance		
		0.01	0.001	0.0001
3	8	6	8	8
4	16	7	14	16
5	32	11	24	32
6	64	15	32	58
7	116	18	46	101
8	202	20	60	163
9	324	20	83	251
10	505	23	103	374
11	753	20	128	531
12	1069	18	157	735
13	1485	14	164	992
14	2006	8	174	1278
15	2628	5	171	1595

Table: Summary of Generated Dataset Configurations

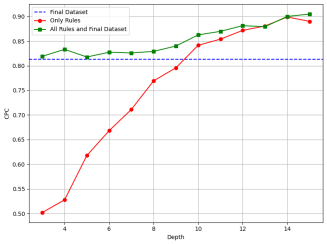
Dataset Type	Description	Count	DT Depths
Final Dataset	Baseline dataset used for evaluation	1	-
Rule-Only Datasets	Extracted rules only	12	3–15
Rules + Final Dataset	Rules and baseline dataset	12	3–15
Rules (Variance 0.01) + Final	Rules filtered (variance ≤ 0.01) and combined with baseline	12	3–15
Rules (Variance 0.001) + Final	Rules filtered (variance ≤ 0.001) and combined with baseline	12	3–15
Rules (Variance 0.0001) + Final	Rules filtered (variance ≤ 0.0001) and combined with baseline	12	3–15


Final Dataset
Only Rules
All Rules and Final Dataset
Mean Absolute Error (MAE)
Depth

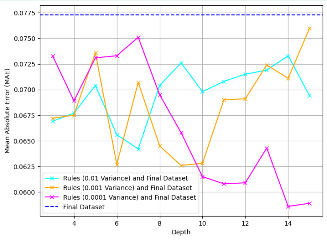
MAE


Final Dataset
Only Rules
All Rules and Final Dataset
R-squared (R²)
Depth

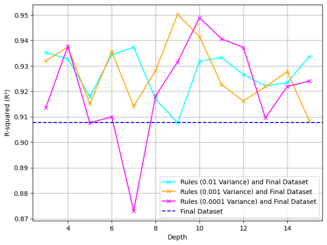
R^2


Final Dataset
Only Rules
All Rules and Final Dataset
CPC
Depth

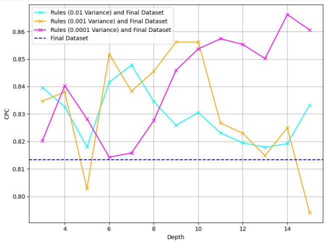
CPC


Mean Absolute Error (MAE)
0.0775
0.0750
0.0725
0.0700
0.0675
0.0650
0.0625
0.0600
4
6
8
10
12
14
Depth
Rules (0.01 Variance) and Final Dataset
Rules (0.001 Variance) and Final Dataset
Rules (0.0001 Variance) and Final Dataset
Final Dataset

MAE


R-squared (R²)
0.95
0.94
0.93
0.92
0.91
0.90
0.89
0.88
0.87
4
6
8
10
12
14
Depth
Rules (0.01 Variance) and Final Dataset
Rules (0.001 Variance) and Final Dataset
Rules (0.0001 Variance) and Final Dataset
Final Dataset

R^2


CPC
0.86
0.85
0.84
0.83
0.82
0.81
0.80
4
6
8
10
12
14
Depth
Rules (0.01 Variance) and Final Dataset
Rules (0.001 Variance) and Final Dataset
Rules (0.0001 Variance) and Final Dataset
Final Dataset

CPC

- Proposed Hybrid Framework:
  - Integrated DT based symbolic rules with NNs for travel demand prediction.
  - Combined interpretability of DT rules with the non-linear learning power of NNs.
- Future Work:
  - **Dynamic rule selection**: Moving away from static predefined variance thresholds to more flexible approaches.
  - **Exploring relevant dataset**: Including the features like weather which highly influenced the trip demand.

C. Wei, T. Lu, and X. Yan, "A logistic regression model with a hierarchical random error term for analyzing the utilization of public transport," *Mathematical Problems in Engineering*, vol. 2015, no. 1, p. 430926, 2015.

G. Guo and T. Zhang, "A residual spatio-temporal architecture for travel demand forecasting," *Transportation Research Part C: Emerging Technologies*, vol. 115, p. 102639, 2020.

S. Hu, C. Xiong, P. Chen, and P. Schonfeld, "Examining nonlinearity in population inflow estimation using big data: An empirical comparison of explainable machine learning models," *Transportation Research Part A: Policy and Practice*, vol. 174, p. 103743, 2023.

E.-J. Kim, "Analysis of travel mode choice in seoul using an interpretable machine learning approach," *Journal of Advanced Transportation*, vol. 2021, no. 1, p. 6685004, 2021.

M. T. Kashifi, A. Jamal, M. S. Kashefi, M. Almoshaogeh, and S. M. Rahman, "Predicting the travel mode choice with interpretable machine learning techniques: A comparative study," *Travel Behaviour and Society*, vol. 29, pp. 279–296, 2022.

Thank You