Graduate Students, Social Sciences and Humanities

An exPlainable machine learning model for Blue-collar workers' travel behavior modeling : The case of Qatar

Aya Hasan Alkhereibi, Ali Abuzaid, Tadesse Wakjira

1. INTRODUCTION

RESEARCH PROBLEM

- Statistics show that car driving is the dominant transport mode amongst Qatari citizens and residents, which results in a non-sustainable transportation system.
- Blue-collar worker represents the biggest population segment, thus investigating their travel mode choice may substantially contribute to the country transport system sustainability.

AIM & OBJECTIVES

- This research provides comprehensive modeling of the blue-collar travel diary using ML models to predict travel mode choice preferences and their behavioral aspect by examining the most associated variables

RESEARCH QUESTIONS

- What are the main characteristics of blue-collar mode choice in the state of Qatar.
- How does machine learning models perform for mode choice prediction.
- What are the main parameters associated with mode choice.

RESEARCH MOTIVATION

- The state of Qatar has a diverse population, the majority of the population is from the blue-collar segment (79.7% -89.1%)
- Population composition, weather, and lifestyle in Qatar promote its uniqueness, putting to the table the need for significance investigations of the blue collar segment mode choice for the better operationalizing of transit systems.

MACHINE LEARNING MODELS

Naïve Bayes (NB) — Naïve Bayes assumes that each attribute variable is independent, and it considers all attributes contribute equally to the outcome. It was considered in this study as a baseline classifier for the dataset.

Support Vector Machine (SVM) — SVM with a radial basis kernel was employed . The cost constraint violation is set to 1, and the gamma parameter for the kernel is set to 0.003 with 3398 support vectors.

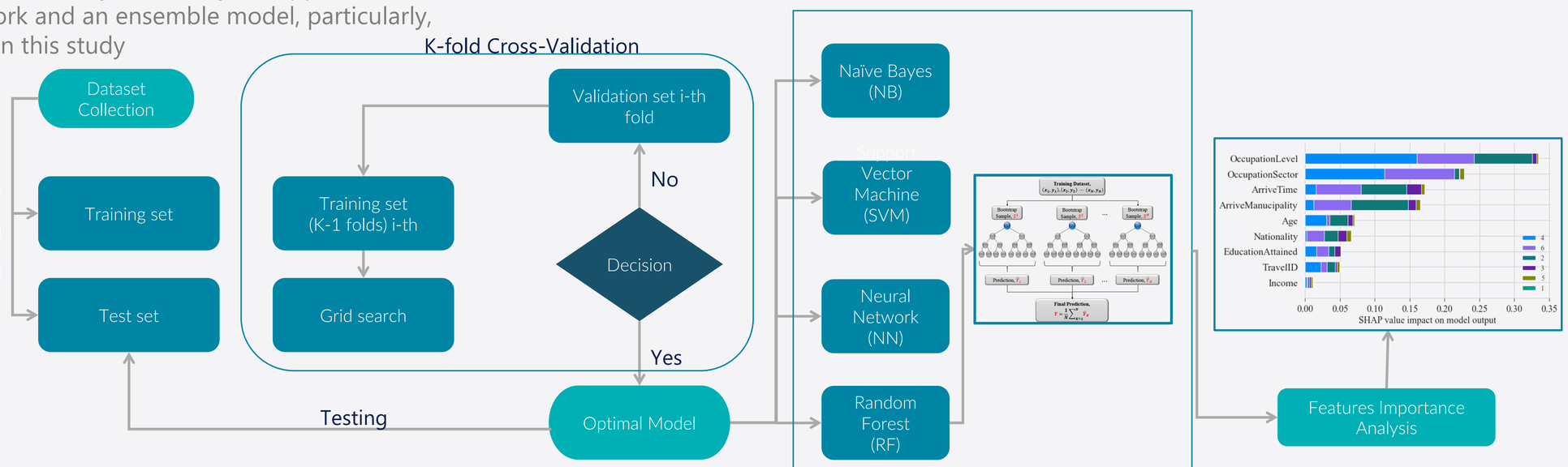
Neural Network (NN) — An NN model with a single hidden layer of 3 units is used. The connection weights are trained by backpropagation with a weight decay constant of 0.1.

Random Forest (RF) — In the RF model (a) generate N bootstrap samples randomly from the training set , (b) grow N number of decision trees (c) determine the prediction from each tree, and (d) compute the final prediction as the mean of the prediction from T trees

2. METHODS & MATERIALS

METHODOLOGY AND DATA COLLECTION

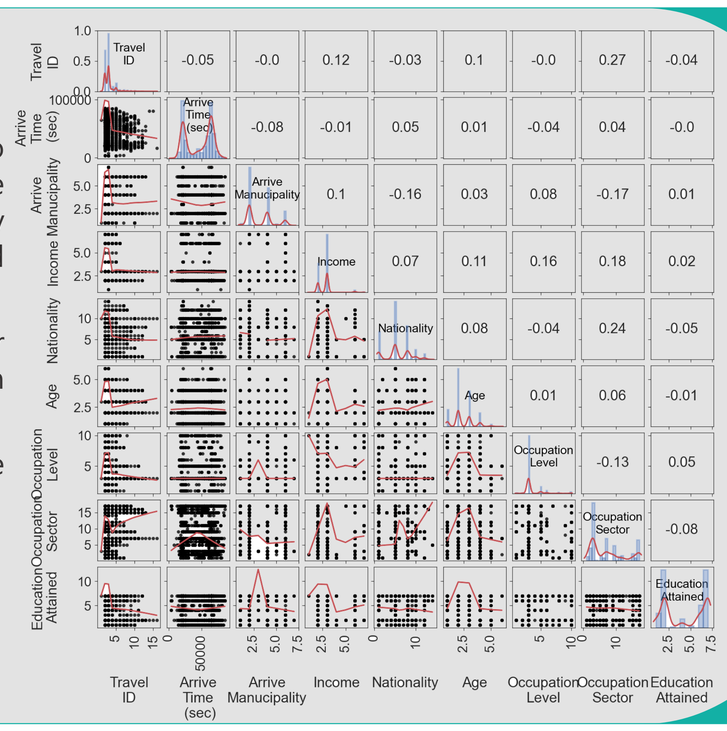
Three different single models, namely, Naïve Bayes, support vector machine, and Neural Network and an ensemble model, particularly, random forest are used to in this study



3. KEY FINDINGS

DATA OVERVIEW

- The primary data used to test the four ML models are obtained from the Ministry of Transportation and Communication's (MoTC)
- survey of blue-collar worker travel dairy, conducted in 2018,
- 1051 blue-collar were interviewed.



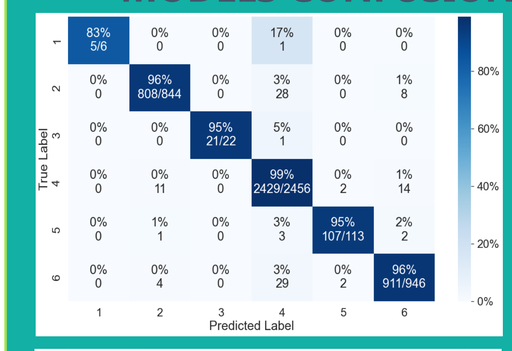
MODELS PERFORMANCE

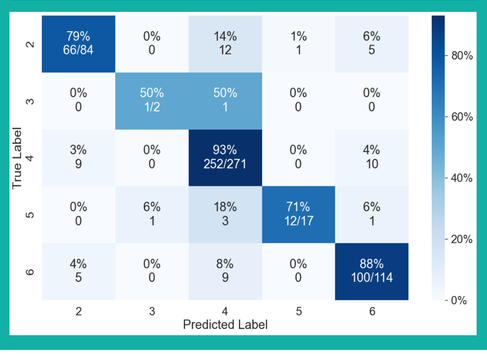
- The accuracy of each model segregated by classes of travel mode choice The proposed RF model showed the highest predictive performance followed by NN, NB, and SVM models respectively.
- Moreover, a consistent rank of ML models is obtained based on its robust measure of reliability kappa.

	Bike	Vehicle	Private Shuttle	Public Transport	Walk	Other	Prediction Accuracy	Kappa
Naïve Bayes (NB)	0.999	0.726	0.791	0.800	0.828	0.832	0.752	0.582
Support Vector Machine (SVM)	0.500	0.614	0.729	0.500	0.778	0.828	0.697	0.445
Neural Network (NN)	0.500	0.810	0.842	0.695	0.889	0.828	0.833	0.696
Random Forest (RF)	0.917	0.963	0.966	0.930	0.968	0.958	0.97	0.937

4. CONCLUSION

MODELS CONFUSION MATERIX





The performance of the ML model can also be investigated with the help of a confusion matrix, which is a table presenting the observed mode choice versus the predicted mode choice. Precision refers to the percentage of the predicted mode choice that the ML model correctly assigns.

CONCLUSION & RECOMMENDATIONS

- Mainly, this study aimed to model the mode choice for blue-collar workers' travel behavior considering several aspects related to their travel and sociodemographic characteristics, such as income, education, occupation, and nationality.
- For that purpose, the authors utilized the ML models to capture blue-collar mode choice based on the cross-validation to capture the models' capabilities of prediction.

Recommendation for Future Work

This study investigated the ML models for blue-collar workers' travel behavior from a predictive aspect, yet more work could be done in terms of

- behavioral insights of their travel behavior, thus more studies could be obtained to deliver a realistic mode choice modeling for blue-collar concerning their behavioral outputs.
- Furthermore, an investigation can be conducted to study the effect of balancing the data on the obtained results.