Bayesian Network Analysis for Car Evaluation Dataset

Raghda Al-taei

2024

1 Introduction

In this report, we analyze the Car Evaluation dataset using Bayesian Networks. The aim is to evaluate different Bayesian network structures and compare their accuracy using 10-fold cross-validation.

2 Datasets

The Car Evaluation dataset is used for this analysis. It contains the following attributes:

- **buying**: Buying price (values: vhigh, high, med, low)
- **maint**: Maintenance cost (values: vhigh, high, med, low)
- **doors**: Number of doors (values: 2, 3, 4, 5more)
- **persons**: Capacity in terms of persons to carry (values: 2, 4, more)
- **lug_boot**: Size of luggage boot (values: small, med, big)
- **safety**: Estimated safety of the car (values: low, med, high)
- **class**: Classification of the car (values: unacc, acc, good, vgood)

The target variable is **class**, and the goal is to build Bayesian Networks that can predict this class based on the other attributes.

3 Data Processing

The dataset is pre-processed before being fed into the Bayesian Network models. Key steps include:

- Converting categorical values into numerical labels.

- Splitting the dataset into 10 folds for cross-validation.
- Preparing training and test sets for each fold.

The data is then used to train three different Bayesian Network structures, whose accuracy is compared using 10-fold cross-validation.

4 Part A: Bayesian Network Structures

In this part, we construct three different Bayesian Network structures and evaluate their accuracy. The details of each structure are as follows:

4.1 Structure 1

Edges: [('class', 'buying'), ('buying', 'maint'), ('class', 'doors'), ('doors', 'persons'), ('lug_boot', 'class'), ('buying', 'safety')]

Explanation: This structure assumes that the class variable directly influences factors like buying, doors, and lug_boot, while buying affects both maint and safety. The relationship between doors and persons suggests that the number of doors is directly related to seating capacity.

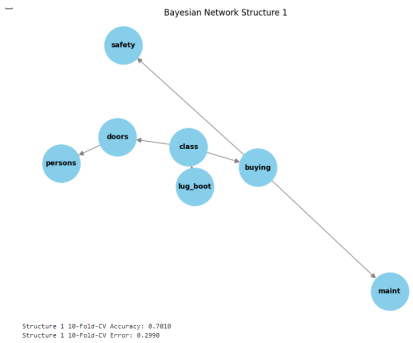

Figure 1: Bayesian Network Structure 1

4.2 Structure 2

Edges: [('class', 'buying'), ('buying', 'maint'), ('doors', 'persons'), ('lug_boot', 'maint'), ('safety', 'class')]

Explanation: This structure suggests that safety has a direct impact on class, while buying and maint are closely related. It maintains the relationship between doors and persons, reflecting how the number of doors is related to seating capacity. The link between lug_boot and maint reflects possible design impacts.

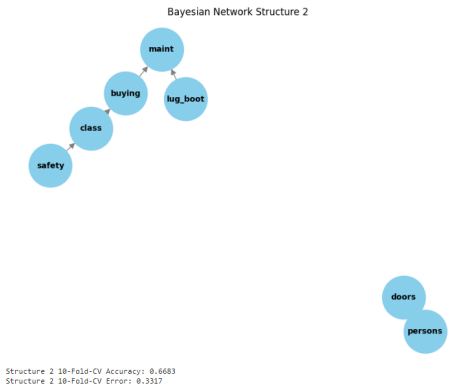

Figure 2: Bayesian Network Structure 2

4.3 Structure 3

Edges: [('class', 'buying'), ('buying', 'maint'), ('maint', 'doors'), ('class', 'safety'), ('persons', 'class')]

Explanation: In this structure, class influences both buying and safety, highlighting the role of car class in determining its price and safety standards. The relationship between maint and doors indicates that maintenance may depend on car design, while persons directly influencing class suggest a connection between capacity and classification.

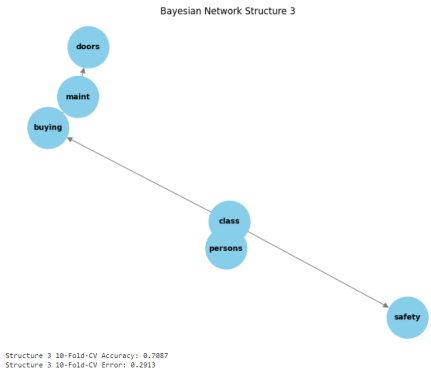

Figure 3: Bayesian Network Structure 3

5 Accuracy of Each Structure

This section presents the accuracy for each fold of the three Bayesian Network structures.

5.1 Structure 1 Accuracy

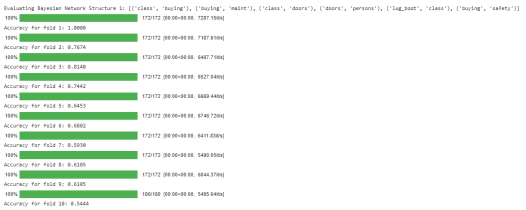

Figure 4: Fold Accuracies for Structure 1

5.2 Structure 2 Accuracy

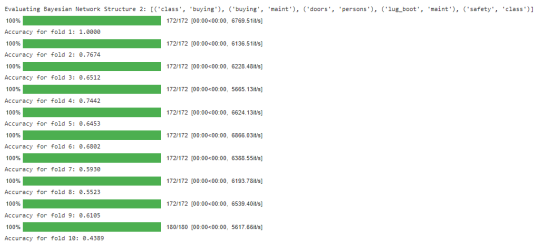

Figure 5: Fold Accuracies for Structure 2

5.3 Structure 3 Accuracy

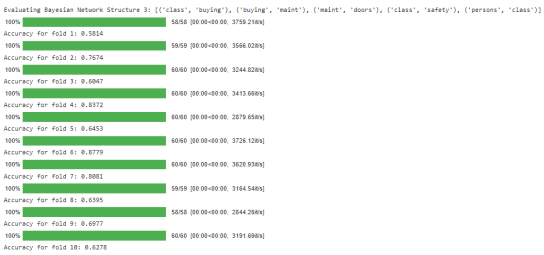

Figure 6: Fold Accuracies for Structure 3

6 Conclusion

The three Bayesian Network structures were evaluated using 10-fold cross-validation. Their accuracy results demonstrate how different network structures

influence the performance of class prediction in the Car Evaluation dataset. By analyzing these structures, we gain insights into the relationships between car attributes and their impact on the classification task.