RANKING APPLICATIONS FOR NURSERY SCHOOLS CLASSIFICATION PROBLEM

Viviane Adohouannon[1], Kate Alexander[1], Juan Arangote[1], Dian Azbel[1], & Igor Baranov[1]

[1] York University School of Continious Studies

Author Note

Group TFG, CSDA1010SUMA18, Lab 1 - Classification Problem

Abstract

Nursery Database was derived from a hierarchical decision model originally developed to rank applications for nursery schools

RANKING APPLICATIONS FOR NURSERY SCHOOLS CLASSIFICATION PROBLEM

Introduction

Background

Objective and Hypothesis

Assessment of Situation

Plan

Data understanding

Preparation

Table 1

Nursery Data Header

parents	has_nurs	form	children	housing	finance	social	health	class
usual	proper	complete	1	convenient	convenient	nonprob	recommended	recommend
usual	proper	complete	1	convenient	convenient	nonprob	priority	priority
usual	proper	complete	1	convenient	convenient	nonprob	not_recom	not_recom
usual	proper	complete	1	convenient	convenient	slightly_prob	recommended	recommend
usual	proper	complete	1	convenient	convenient	slightly_prob	priority	priority
usual	proper	complete	1	convenient	convenient	slightly_prob	not_recom	not_recom
usual	proper	complete	1	convenient	convenient	problematic	recommended	priority
usual	proper	complete	1	convenient	convenient	problematic	priority	priority
usual	proper	complete	1	convenient	convenient	problematic	not_recom	not_recom
usual	proper	complete	1	convenient	inconv	nonprob	recommended	very_recom
usual	proper	complete	1	convenient	inconv	nonprob	priority	priority
usual	proper	complete	1	convenient	inconv	nonprob	not_recom	not_recom
usual	proper	complete	1	convenient	inconv	slightly_prob	recommended	very_recom
usual	proper	complete	1	convenient	inconv	slightly_prob	priority	priority
usual	proper	complete	1	convenient	inconv	slightly_prob	not_recom	not_recom

Note. This is a nursery data header

parents has_nurs form children

great_pret :4320 critical :2592 complete :3240 1 :3240

pretentious:4320 improper :2592 completed :3240 2 :3240

usual :4320 less_proper:2592 foster :3240 3 :3240

proper :2592 incomplete:3240 more:3240

very_crit :2592

housing finance social

convenient:4320 convenient:6480 nonprob :4320

critical :4320 inconv :6480 problematic :4320

less_conv :4320 slightly_prob:4320

health class

not_recom :4320 not_recom :4320

priority :4320 priority :4266

recommended:4320 recommend : 2

spec_prior:4044

very_recom: 328

‘‘‘

Cleaning

Feature engineering

Modeling

Splitting the dataset into train and test

Decision Tree - Iteration 1

Decision Tree model fit.

Decision Tree model evaluation.

Random Forest - Iteration 2

Random Forest model fit.

Random Forest model evaluation.

Random Forest - Iteration 3

Random Forest model fit.

Random Forest model evaluation.

Dicsussion

Conclusion

References