

HackRcity 2019

Model predykcji popularności turystycznej miejscowości

czterej_panceRni:

Bartek Karaban, Patryk Wielopolski, Daria Puchalska, Jarek Opała

Zadanie

Explore, Predict & Explain

Workflow analizy

- 1. Przygotowanie danych**
- 2. Budowa modeli**
- 3. Analiza wybranego modelu**
- 4. Wyjaśnienie**

1. Przygotowanie danych

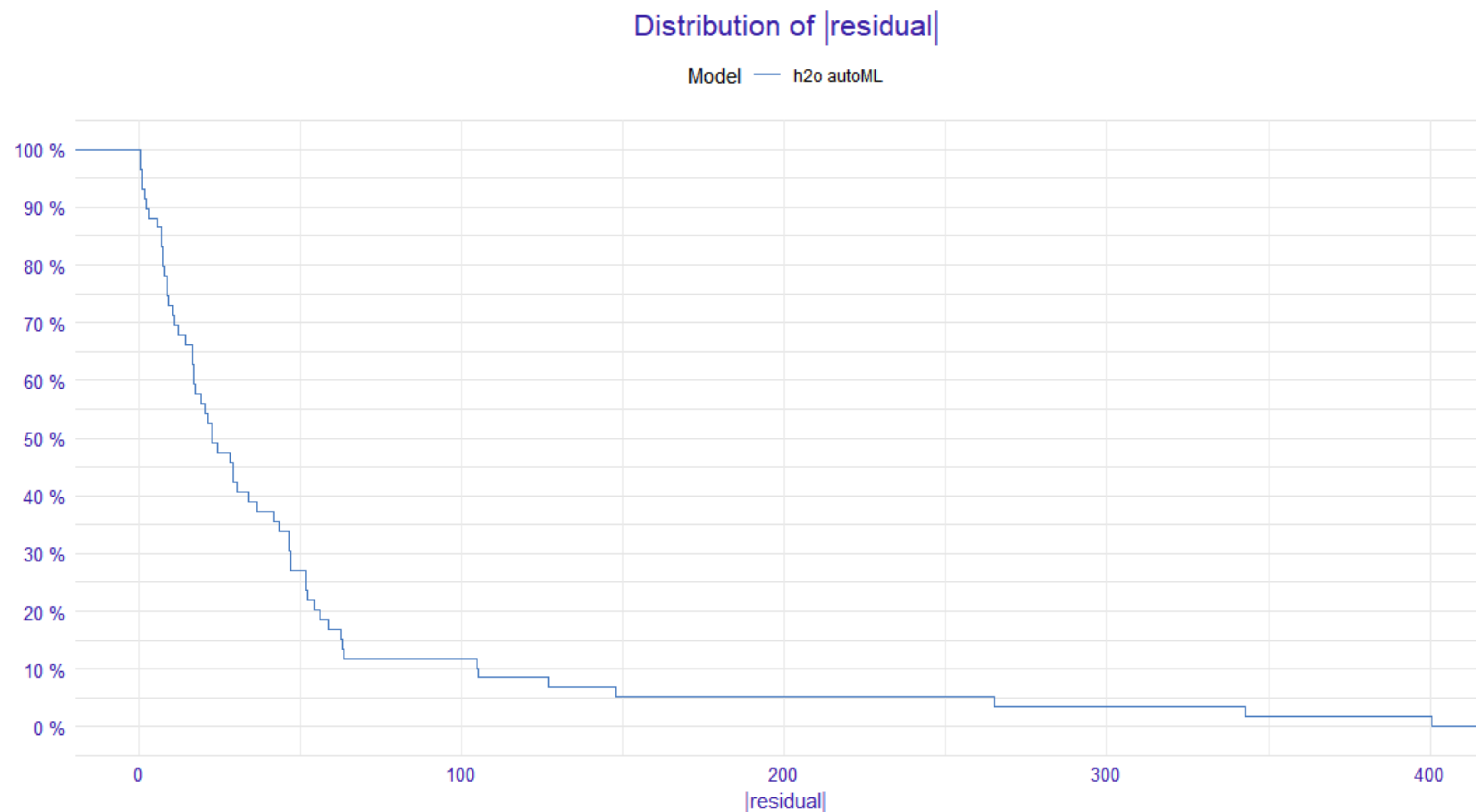
1. Wzbogacanie danych z innych źródeł,
2. Usuwanie zmiennych z dużą liczbą NA's oraz 1 unikalną wartością
3. Selekcja zmiennych w oparciu o kryterium korelacji

2. Przygotowanie modeli

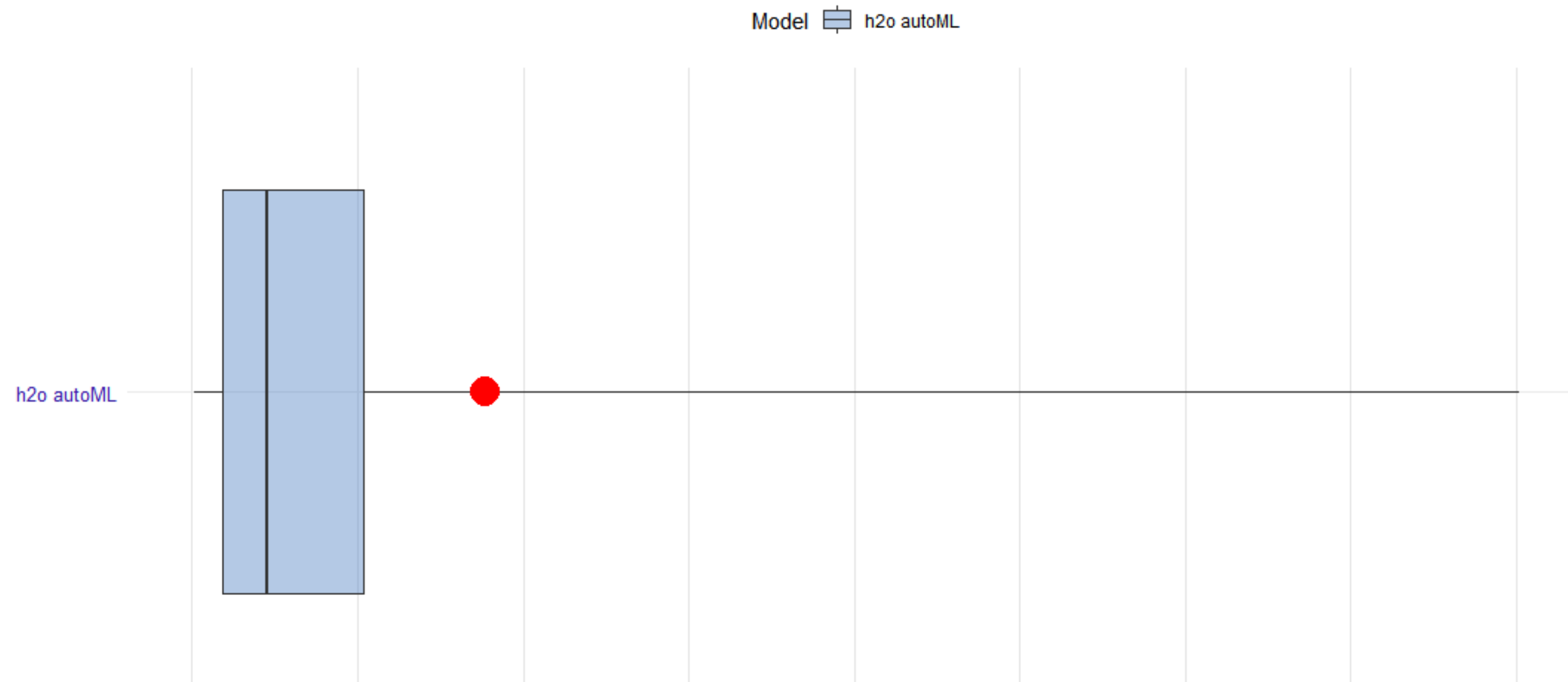
1. Wybraliśmy framework h2o
2. Zastosowaliśmy podejście autoML
3. Spośród rankingu modeli wybraliśmy model Deep Learningowy.
4. Kryterium wyboru był najmniejszy błąd MAE

| | model_id | mean_residual_deviance | rmse | mse | mae |
|---|--|------------------------|-----------|------------|-----------|
| 1 | DeepLearning_grid_1_AutoML_20191116_183450_model_1 | 9401.813 | 96.96295 | 9401.813 | 54.04622 |
| 2 | DeepLearning_grid_1_AutoML_20191116_183450_model_2 | 26637.509 | 163.21001 | 26637.509 | 126.13250 |
| 3 | GBM_grid_1_AutoML_20191116_183450_model_36 | 48231.457 | 219.61661 | 48231.457 | 173.25309 |
| 4 | DeepLearning_grid_1_AutoML_20191116_183450_model_4 | 57929.748 | 240.68600 | 57929.748 | 118.45586 |
| 5 | GLM_grid_1_AutoML_20191116_183450_model_1 | 104224.233 | 322.83778 | 104224.233 | 159.72276 |
| 6 | DeepLearning_grid_1_AutoML_20191116_183450_model_3 | 113769.862 | 337.29788 | 113769.862 | 219.98307 |

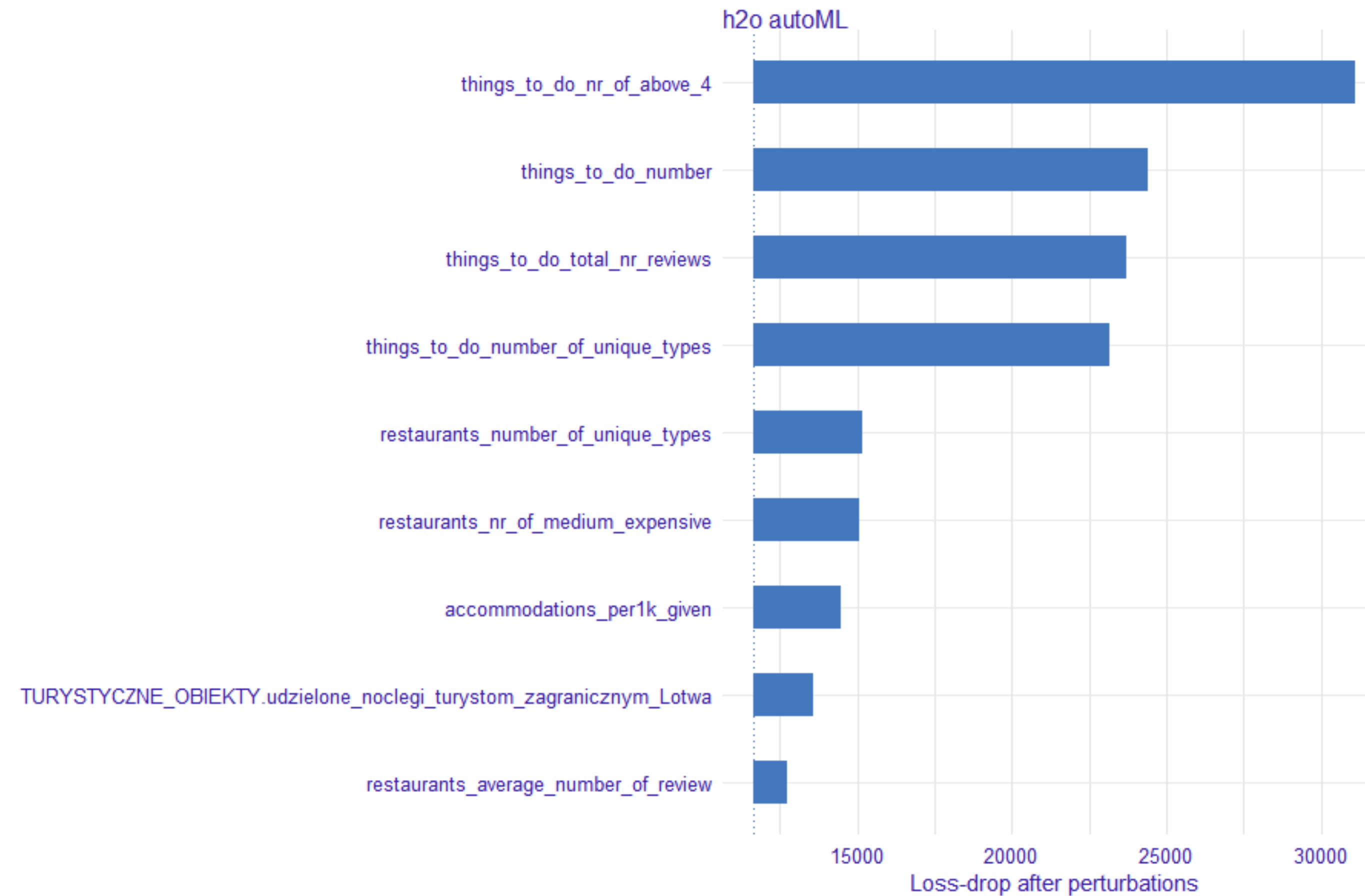
3. Analiza wybranego modelu



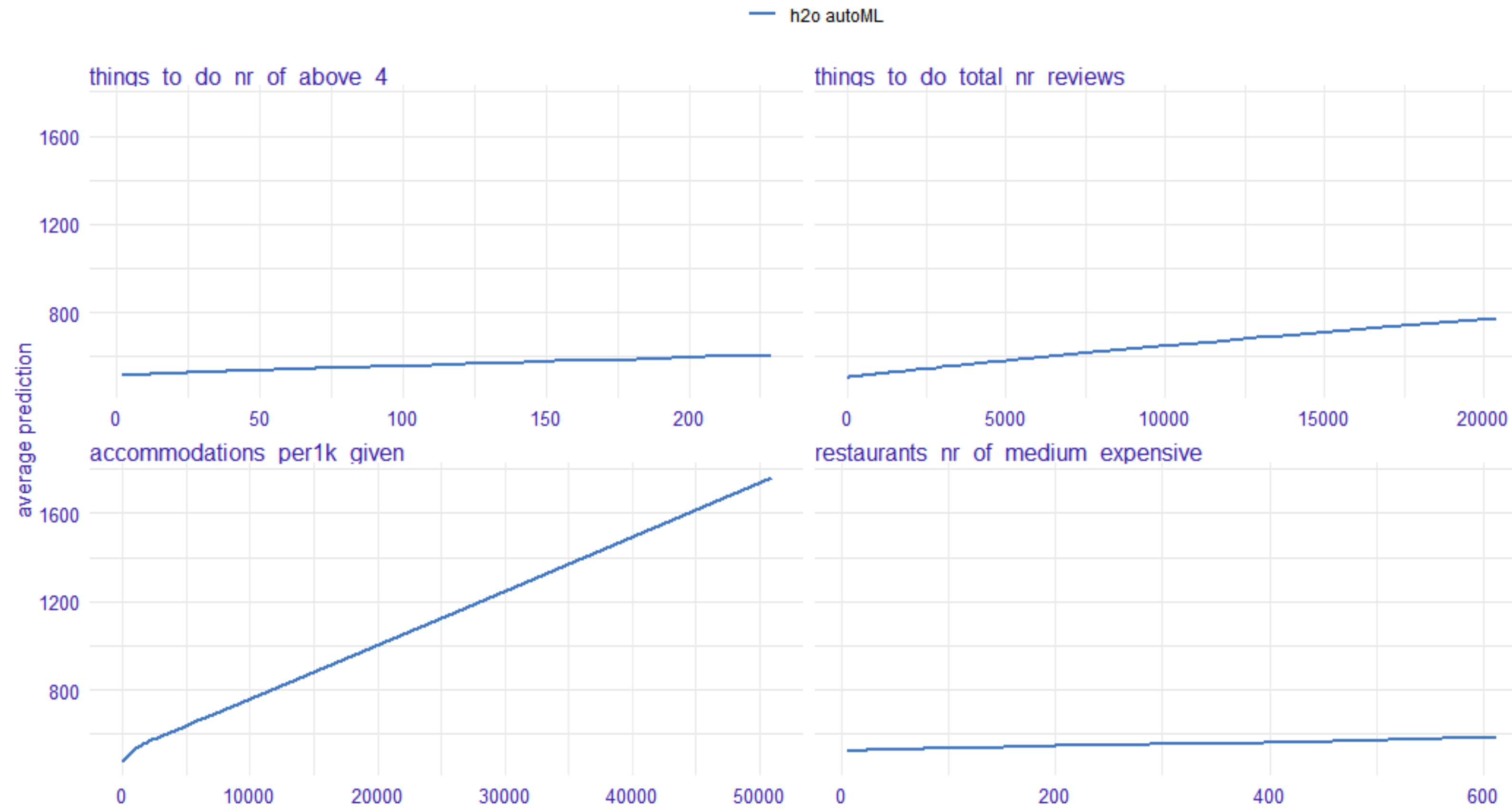
3. Analiza wybranego modelu



4. Wyjaśnienie modelu



4. Wyjaśnienie modelu



Zadanie Optimize

Metoda rozwiązania

- Analiza preferencji rodziny.
- Wybranie reprezentatywnych zmiennych.
- Stworzenie statystyki atrakcyjności miast.
- Znalezienie optymalnej trasy.

Proponowana trasa

