

Universidade Federal do Rio Grande do Sul
Instituto de Informática



INF01017
Aprendizado de Máquina

Atividade Prática 06

Descrição dos Dados e Análise de Necessidades de Pré-processamento

Luís Filipe Martini Gastmann (00276150)
Pedro Lubaszewski Lima (00341810)
Vinícius Boff Alves (00335551)

Turma U

9 de novembro de 2024

Sumário

1.1	Definição do Problema e Coleta de Dados	2
2.1	Análise Exploratória e Pré-processamento dos Dados	3

1.1 Definição do Problema e Coleta de Dados

O objetivo deste trabalho é prever o consumo médio de combustível de um carro através de algumas das suas características e origens de fabricação. Alguns atributos das instâncias são a marca, a quantidade de cilindros, o porte do veículo etc.

O conjunto de dados utilizado para desenvolver este trabalho foi obtido da seguinte página do Kaggle: Explore Car Performance: Fuel Efficiency Data. Essa tarefa contará com diversas técnicas de preparação dos dados para posteriormente iniciar a seleção e avaliação de modelos para essa tarefa.

2.1 Análise Exploratória e Pré-processamento dos Dados

Este *dataset* possui 550 instâncias, com 11 atributos preditores e 1 atributo alvo. Esse último torna a tarefa dos modelos em regressão, visto que o objetivo aqui é prever o consumo médio de combustível de carros.

Dos atributos preditores, observa-se que há 5 atributos numéricos (*city mpg*, *cylinders*, *displacement*, *highway mpg* e *year*). Além deles, os 5 atributos restantes são categóricos: *class*, *drive*, *fuel type*, *make*, *model* e *transmission*.