

# Seleção de Características

## Abordagem Filtro

Prof. Dr. Danilo S. Sanches



# Introdução

- Seleção de atributos é o processo que remove atributos irrelevantes e redundantes do conjunto de dados;
- O modelo, por sua vez, será de complexidade reduzida e, portanto, mais fácil de interpretar.

All Features



Feature Selection



Final Features

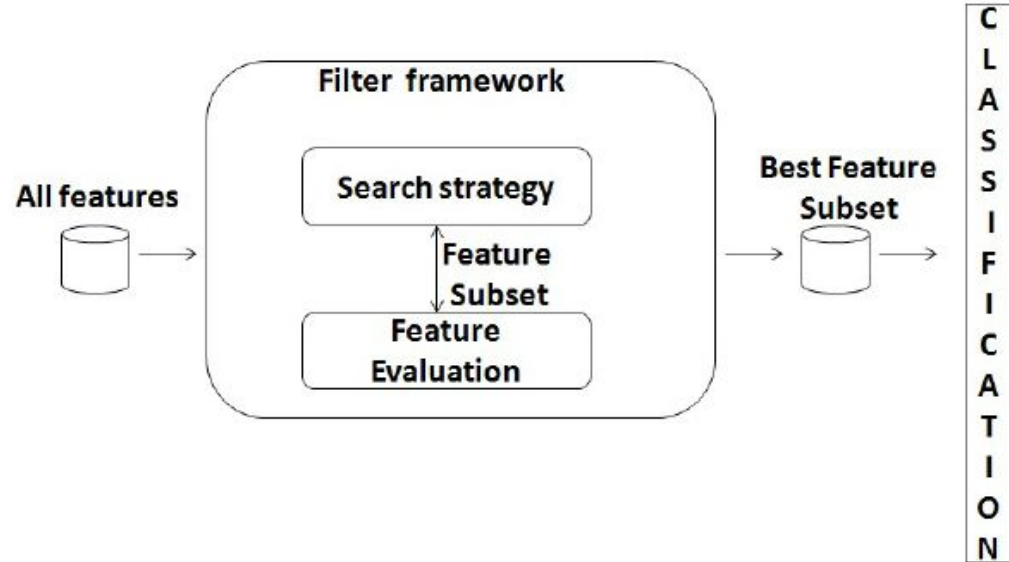


Fonte:

<https://themanoftalent.medium.com/feature-selection-9b1609f1f6b0>

# Abordagem Filtro

- Um subconjunto de atributos é selecionado com base em seu relacionamento com o atributo alvo (classe);
- A seleção não depende de nenhum algoritmo de aprendizado de máquina;
- Pelo contrário, os métodos de filtro medem a “relevância” dos atributos com a saída por meio de testes estatísticos.



# Principais métodos estatísticos

- Pearson's Correlation
- Linear Discriminant Analysis (LDA)
- ANOVA
- Chi squared

X \ Y	Continuous	Categorical
Continuous	Pearson's Correlation	LDA
Categorical	Anova	Chi-Squared

# Métodos estatísticos em Python



## `sklearn.feature_selection`

Feature selection algorithms.

These include univariate filter selection methods and the recursive feature elimination algorithm.

**User guide.** See the [Feature selection](#) section for further details.

# Métodos estadísticos: Scikit-learn

**Biblioteca:** *sklearn.feature\_selection*

f\_classif

ANOVA F-value between label/feature for classification tasks.

mutual\_info\_classif

Mutual information for a discrete target.

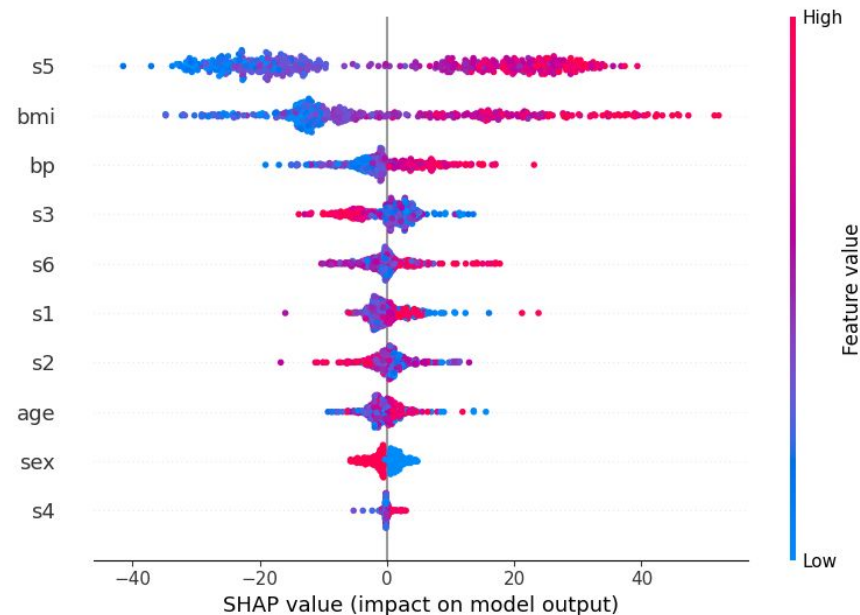
chi2

Chi-squared stats of non-negative features for classification tasks.

# Importância dos atributos



- SHAP Values (***SHapley Additive exPlanations***) é uma abordagem para explicar as previsões feitas por modelos de machine learning;
- Destaca como a presença de cada característica contribui para a previsão de um determinado modelo;
- São baseados na teoria dos jogos;



<https://shap.readthedocs.io/en/latest/index.html>