



AUTOML – PRACA DOMOWA 2

SIWAK PAWEŁ, 313450

WUJKOWSKI DANIEL, 313472



PODSTAWOWA ANALIZA DANYCH

- 500 zmiennych
- Zbiór treningowy: 2000 obserwacji, testowy: 600
- Wyłącznie dane numeryczne
- Bez brakujących danych
- Brak przewagi jednej z klas



0.933

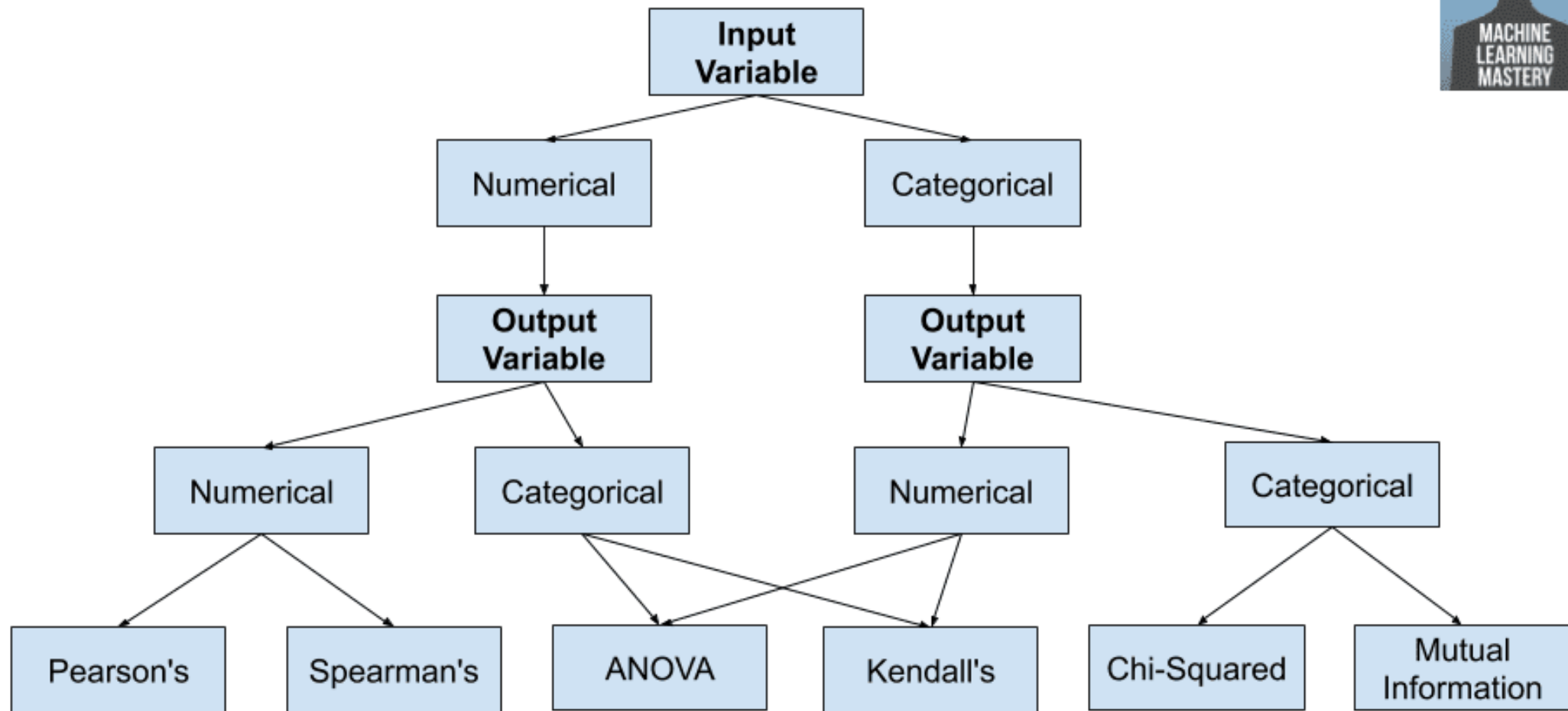
AUTORSKA METODA REDUKCJI LICZBY ZMIENNYCH + SFS

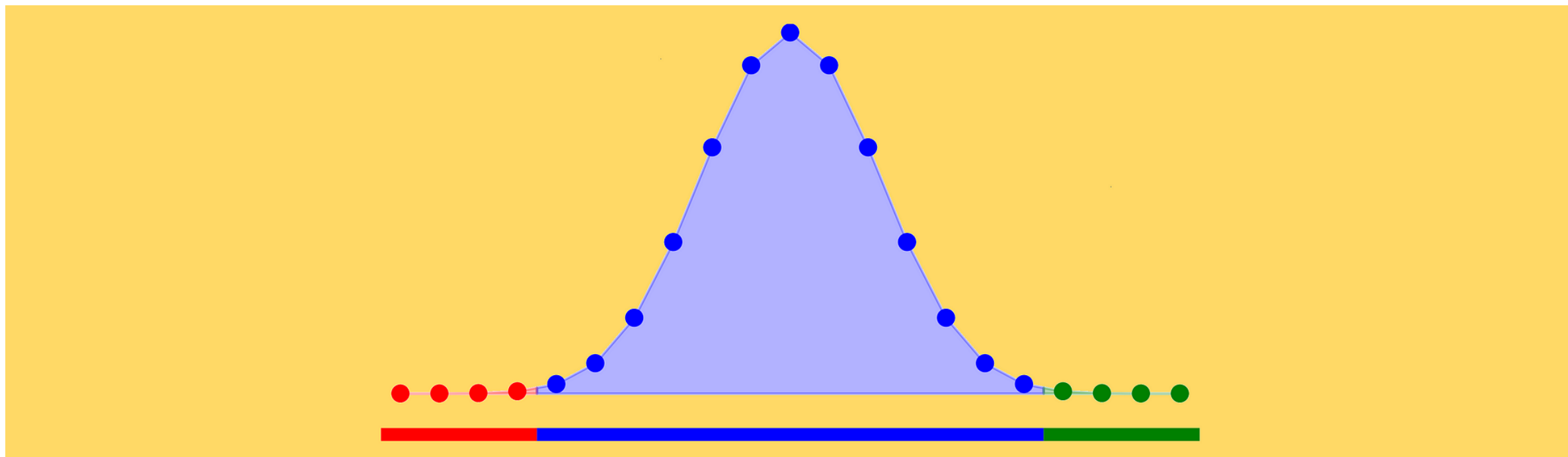
Korelacja
 ≥ 0.001

Mutual Info
 ≥ 0.000001

Chi-kwadrat:
p-value
 ≤ 0.1

How to Choose a Feature Selection Method





BORUTA

KLASYCZNE UCZENIE MASZYNOWE

Rodzaj modelu	Domyślne hiperparametry	Bayes
Random Forest	0.883	0.879
Gradient Boosting	0.762	0.871
SVM	0.827	0.878
Multi Layer Perceptron	0.756	-
Extra Trees	0.886	0.903
Histogram-based Gradient Boosting	0.856	0.873

AUTOML

Framework	Boruta	ANOVA
-----------	--------	-------

TabPFN	0.832	0.856
--------	-------	-------

Autogluon (default preset, 1h)	0.874	0.874
--------------------------------	-------	-------

Autogluon (best preset, 1h)	-	0.898
-----------------------------	---	-------

Autogluon (default preset + stacking + bagging, 1h)	-	0.892
---	---	-------

Autogluon (best preset + stacking + bagging, 1h)	-	0.900
--	---	-------

Autogluon (best preset + stacking + bagging, 8h)	-	0.900
--	---	-------

DZIĘKUJEMY ZA UWAGĘ

- Roberto Battiti. Using mutual information for selecting features in supervised neural net learning. Neural Networks, IEEE Transactions on, 5:537 – 550, 08 1994.
- Miron B. Kursa and Witold R. Rudnicki. Feature selection with the boruta package. Journal of Statistical Software, 36(11):1–13, 2010.
- Jason Brownlee PhD. How to choose a feature selection method for machine learning. Machine Learning Mastery, 2020.
(<https://machinelearningmastery.com/feature-selection-with-real-and-categorical-data/>)