



LTAT.06.010 Pervasive Data Science

Lecture 4

Huber Flores, PhD

ASSOCIATE PROFESSOR

Tartu, Estonia 28/02/2022

Recap

- Each student presented their envisioned testbed



What to do with the data?

- Pre-processing
 - Cleaning the data, adjusting scales, representativeness and validity
- Analysis (**this lecture**)
- Model construction
 - Classical machine learning

Analysis

	Type of Data			
Goal	Measurement (from Gaussian Population)	Rank, Score, or Measurement (from Non-Gaussian Population)	Binomial (Two Possible Outcomes)	Survival Time
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square or Binomial test **	
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel*
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test	Conditional proportional hazards regression*
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression**
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**	Conditional proportional hazards regression**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**	
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*	Cox proportional hazard regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**		Multiple logistic regression*	Cox proportional hazard regression*

[source]

<https://www.graphpad.com/support/faqid/1790/>

PENGUIN: Aquatic Plastic Pollution Sensing using
Autonomous Underwater Vehicles (AUV)s

<https://dl.acm.org/doi/10.1145/3396864.3399704>

PENGUIN: Testbed

RIC



Plastic bottles



Plastic bags



Shampoo bottles



Plastic cards



Six pack rings



Plastic trays



Respiratory masks



Diapers



Food packing



Car panel

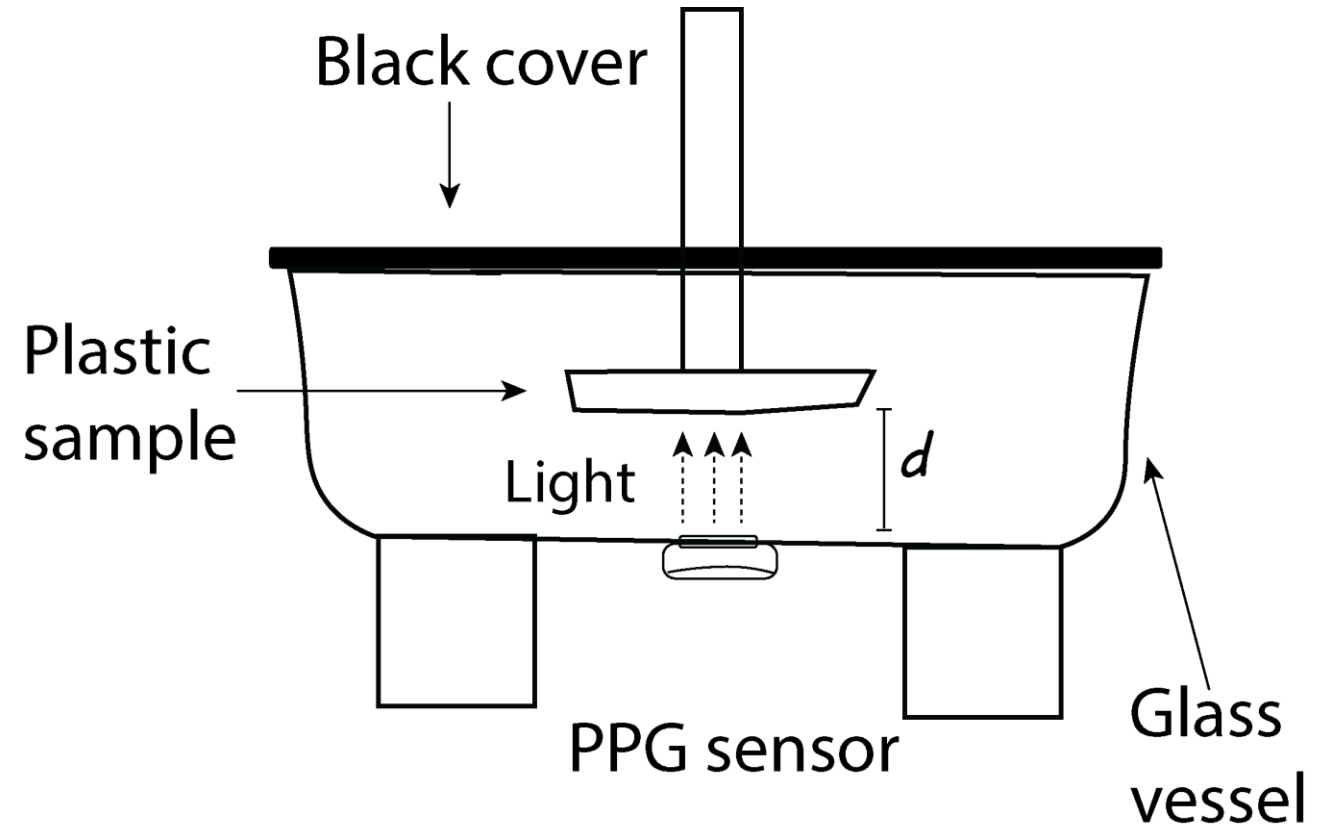
PENGUIN: Testbed



Controlled plastic samples



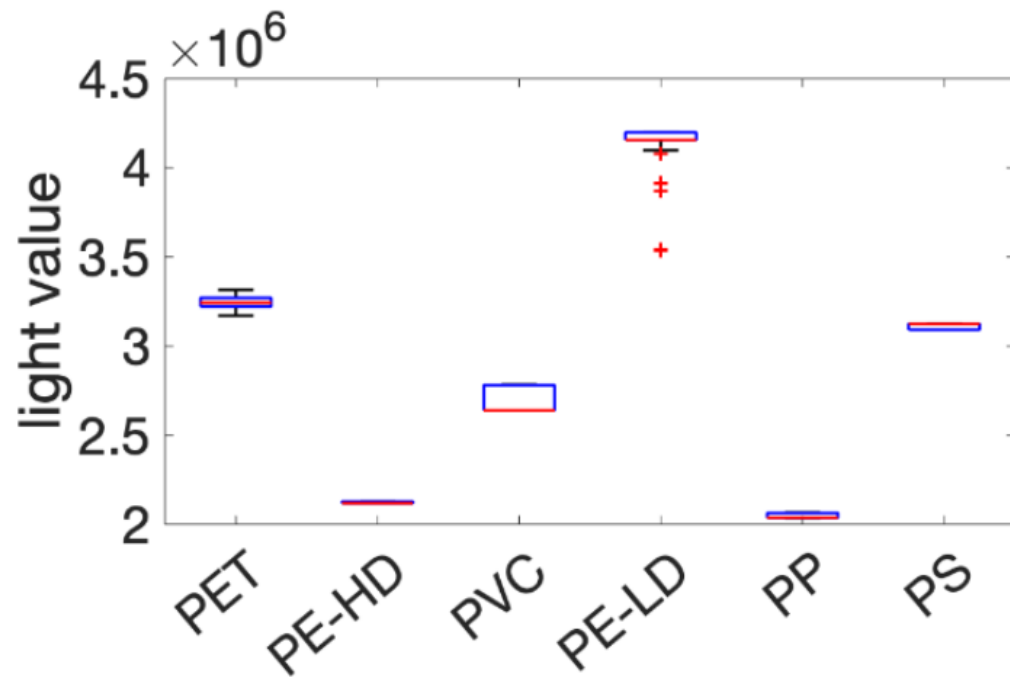
Light sensor through glass



Testbed for plastic recognition underwater

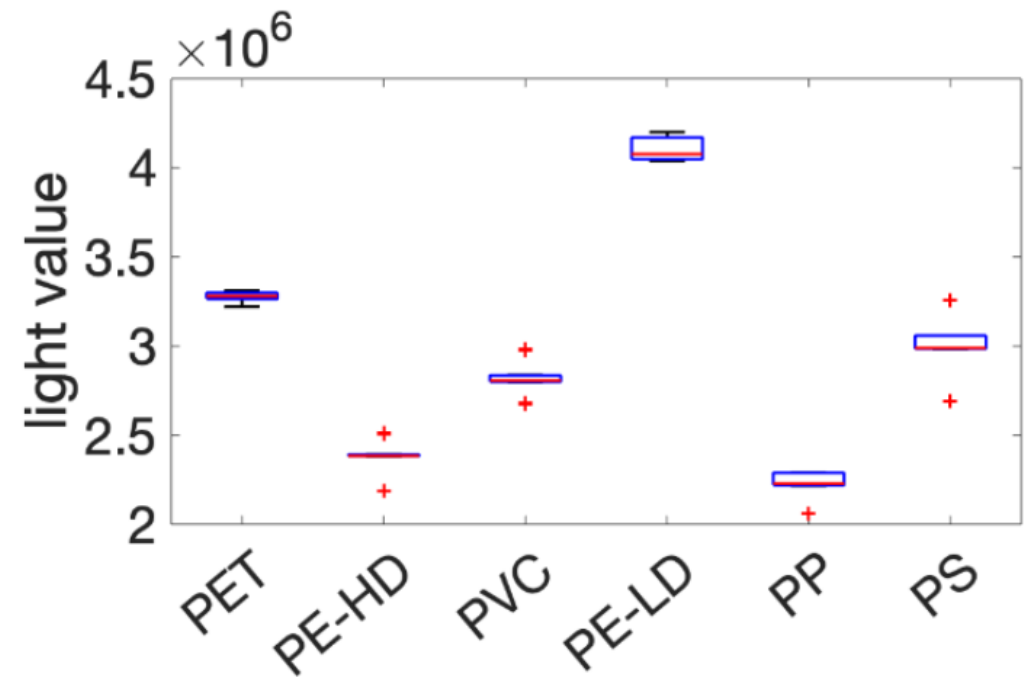
PENGUIN: Evaluation

Result: Green light reflectivity can characterize plastics in different environmental conditions



(a)

Ambient light



(b)

Dark environment

PENGUIN: Evaluation

Statistical analysis in R

```
50 #Effect
51 #Kendall's W uses the Cohen's interpretation guidelines of 0.1 - < 0.3 (small effect), 0.3 - < 0.5
    (moderate effect) and >= 0.5 (large effect).
52 rlight %>% friedman_effsize(fingerprint ~ material |id)
53 # W=0.54, magnitude = large effect
54
55
56
57 # pairwise comparisons
58 #A significant Friedman test can be followed up by pairwise wilcoxon signed-rank tests for identifying
    which groups are different.
59
60 pwc <- rlight %>%
61   wilcox_test(fingerprint ~ material, paired = TRUE, p.adjust.method = "bonferroni")
62 pwc
63
```

PENGUIN: Evaluation

Result: High accuracy to classify different types of plastics underwater

Test	k-NN	Random forest	Average
Cross Validation			
All conditions 6-folds	0.95	0.95	0.95
Ambient 6-folds	0.96	0.95	0.96
Darkness 6-folds	0.94	0.96	0.95
Model data → Predicted			
Ambient → All conditions	0.80	0.80	0.80
Ambient → Darkness	0.69	0.68	0.69
Darkness → All conditions	0.95	0.95	0.95
Darkness → Ambient	0.94	0.92	0.93
Average	89.0	88.7	88.9

Plastic classification accuracy in different experimental conditions.

Model data -> Predicted



Next lecture

Phd students presenting projects (Part I)



Questions?

E-mail: huber.flores@ut.ee