Missing Data Imputation Methods for Food Composition Data

*Gordana Ispirova[1,2], Tome Eftimov[1], Barbara Koroušić Seljak[1]

Objective

Scenario: Nutrient-intake assessment, dietary guidelines
Problem: Missing nutrient data
Where: Nutrition domain
Target: Food composition databases (FCDBs)
Solution: Missing data imputation techniques

FCDBs?

Detailed sets of information on the nutritionally important food components.

Nutrient/Food	Apple	Pear	Peach
Na	?	1	?
K	107	116	190
Mg	5	7	?

Method

Missing data imputation techniques:

1. Fill-in with mean
2. Fill-in with median
3. Non-Negative Matrix Factorization (NMF)
4. Multiple Imputations by Chained Equations (MICE)
5. Nonparametric Missing Value Imputation using Random Forest (MissForest)
6. K-Nearest Neighbors (KNN)

FCDBs are quite incomplete, obtaining complete datasets was not an easy task, and the datasets obtained are small!

Data?

From the national FCDBs of 10 countries, collected by EuroFIR
Nutritional values for foods from four food groups:

- Potassium in Fruits
- Sodium in Fruits
- Sodium in Vegetables
- Protein in Meat

$Nutrient_a$	$Country_1$	$Country_2$	…	$Country_m$
$Food_1$	$Value_{11}$	$Value_{12}$	…	$Value_{1m}$
⋮	⋮	⋮	…	⋮
$Food_n$	$Value_{n1}$	$Value_{n2}$	…	$Value_{nm}$

Results and conclusions

Evaluation:

1. 10%, 20% and 30% of the data is set as missing
2. Calculate the missing values
3. Compare obtained values with actual values

Evaluation criteria:

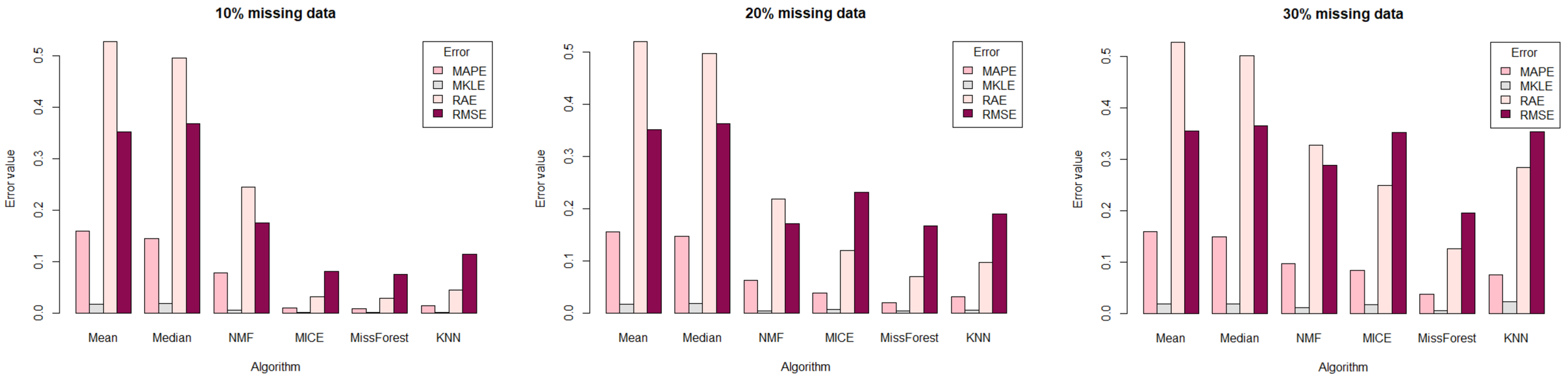
- Mean Absolute Percentage Error (MAPE)
- Mean Kullback-Liebler Divergence Error (MKLDE)
- Relative Absolute Error (RAE)
- Root Mean Square Error (RMSE)

Best technique?

Significantly better results than the classical fill-in with mean or median techniques.

Overall: MICE and Miss Forest yield the best results and deserve further consideration in practice.

The graphs are generated from the evaluation of the dataset with values for potassium in Fruits.



Computer Systems
Jožef Stefan Institute

EuroFIR
European Food Information Resource

ISO Fd Food ERA Chair
1st ISO-FOOD International Symposium on Isotopic and Other Techniques in Food Safety and Quality
April 1-3, 2019, Piran, Slovenia

*gordana.ispirova@ijs.si
[1]Computer Systems Department
Jožef Stefan Institute,
Ljubljana, Slovenia
[2]Jožef Stefan International
Postgraduate School,
Ljubljana, Slovenia