

I. Sala^{1,2}, D. Rozza³, L. Losa³, S. Conti⁴, P. Ferrara³, V. Bagnardi², C. Fornari³, A. Zucchⁱ⁵, R. Ciampichini⁵, G. Sampietro⁵, L.G. Mantovani³



Department of Medicine and Surgery, University of Milano-Bicocca, Milan, Italy
 Department of Statistics and Quantitative Methods, University of Milano-Bicocca, Milan, Italy
 Center for Public Health Research, University of Milano-Bicocca, Monza, Italy
 Occupational Health Service, University of Milano-Bicocca, Milan, Italy
 Epidemiology Service, Health Protection Agency of Bergamo, Bergamo, Italy



EPH23

BACKGROUND

+

costs

economic).

- MATERIALS AND METHODS
- ◆ Population and data source: AHDBs' records of individuals aged ≥18 covered by the Health Protection Agency of Bergamo (Northern Italy), years 2010-2016.
- + Variables of interest: access to inpatient and outpatient services and total annual costs based on healthcare service tariffs.
- A regression tree was trained to predict yearly costs for subjects in 2015 based on their access history to the NHS from 2010 to 2014. Validation over the period 2011 to 2016.²
- Optimal hyperparameters selected using a 10-fold cross-validation approach.
- Objective: to predict direct healthcare costs for the Italian National Healthcare Service (NHS) through the identification of homogeneous population segments for healthcare expenditures.

Predicting future expenditures is crucial for healthcare planning, as

longer life expectancies and increasing prevalence of chronic diseases drive up demand for healthcare services and related

Machine learning (ML) techniques can be applied to administrative

healthcare databases (AHDB), which are a source of large

amounts of inpatient and outpatient data (demographic, clinical,

 For each segment identified by the tree, a prediction error was calculated as the ratio of the difference between predicted and actual total costs and the predicted total cost.

RESULTS

- + In 2010, 70.7% of the 902,023 included subjects used at least one inpatient (n=99,860) and/or outpatient service (n=631,451), for a total cost of €692,298,400.
- High-cost subjects (>€15,000 yearly), accounting for 0.8% of the population, absorbed 28.7% of total costs.
- The figure shows the application of the decision tree rule to the test set for predicting healthcare costs in 2016 based on 2011-2015 data. Some groups were highlighted and described, in order to
 clarify results interpretation.



CONCLUSIONS

- These preliminary results identify several predictors of direct costs for the NHS.
- To achieve more accurate predictions, further ensemble methods, such as random forests, are currently being applied, and will soon give preliminary results.
- The time period will also be extended to include data up to 2022, and pharmaceutical information will be integrated.

REFERENCES

 Rosella, L. C., & Harish, V. (2022). Machine Learning Applied to Routinely Collected Health Administrative Data. Healthcare Quarterly (Toronto, Ont.), 25(1), 12-16.

2. Therneau, T., Atkinson, B., Ripley, B., & Ripley, M. B. (2015). Package 'rpart'. Available online: https://cran.r-project.org/web/packages/rpart/rpart.pdf (accessed on 4 September 2024).

Acknowledgements: Funding provided by the Italian Ministry of University and Research, Next Generation EU Program, and Mediolanum Assicurazioni SpA.