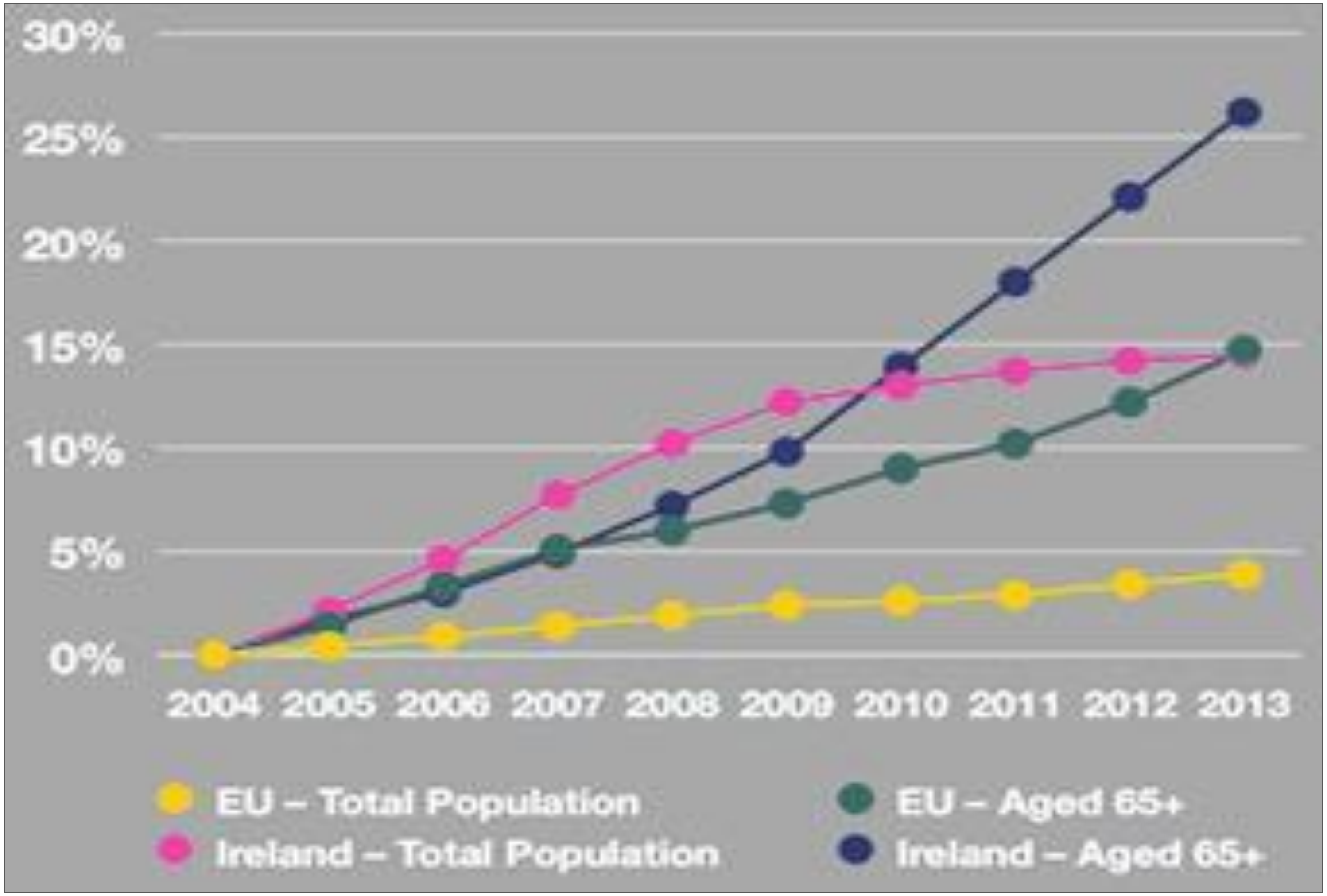
Clustering-Aided Approach for Predicting Patient Outcomes with Application to Elderly Healthcare in Ireland

Mahmoud Elbattah, Owen Molloy
mahmoud.elbattah@nuigalway.ie

National University of Ireland, Galway

Challenge to Healthcare: Population Ageing



Source : Health Service Executive. Annual Report and Financial Statements, 2014.

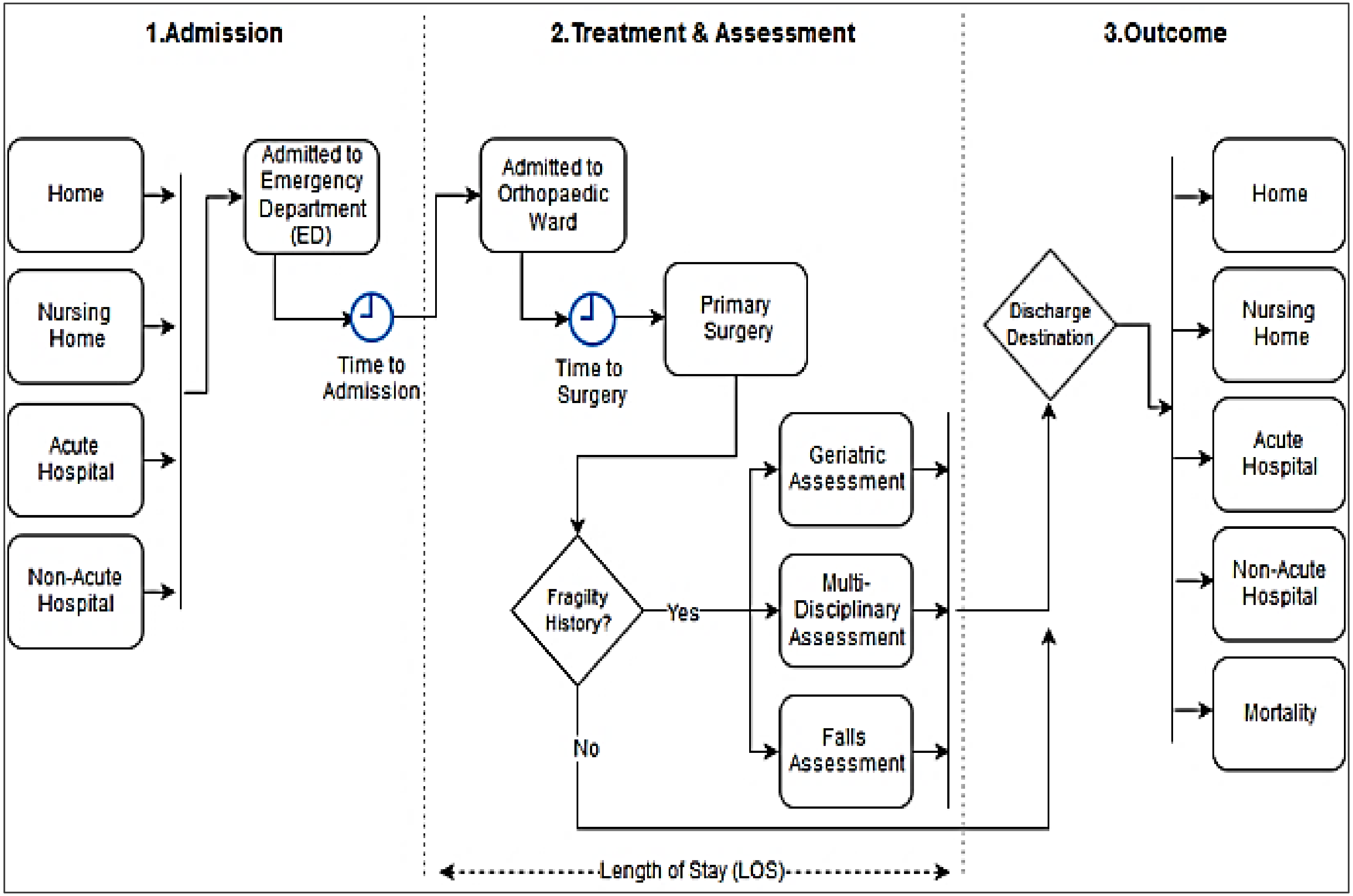
OÉ Gaillimh NUI Galway

Our Focus: Hip Fracture Care in Ireland

- A good exemplar of elderly healthcare.
- Exponentially increasing with age.[1]
- Identified as one of the most serious injuries resulting in lengthy hospital admissions and high costs.[2]
- High quality data available through the Irish Hip Fracture Database (IHFD).

Sources :[1] Gullberg, B., Johnell, O. and Kanis, J.A., 1997. World-wide projections for hip fracture. Osteoporosis international, 7(5), pp.407-413.
[2]http://www.hse.ie/eng/services/publications/olderpeople/Executive_Summary_Strategy_to_Prevent_Falls_and_Fractures_in_Ireland%E2%80%99s_Ageing_Population.pdf

Overview: Hip Fracture Care Scheme



Objectives

- To predict the inpatient length of stay (LOS).
- To predict the patient's discharge destination.

Importance of Predicting LOS and Discharge Destination

LOS:

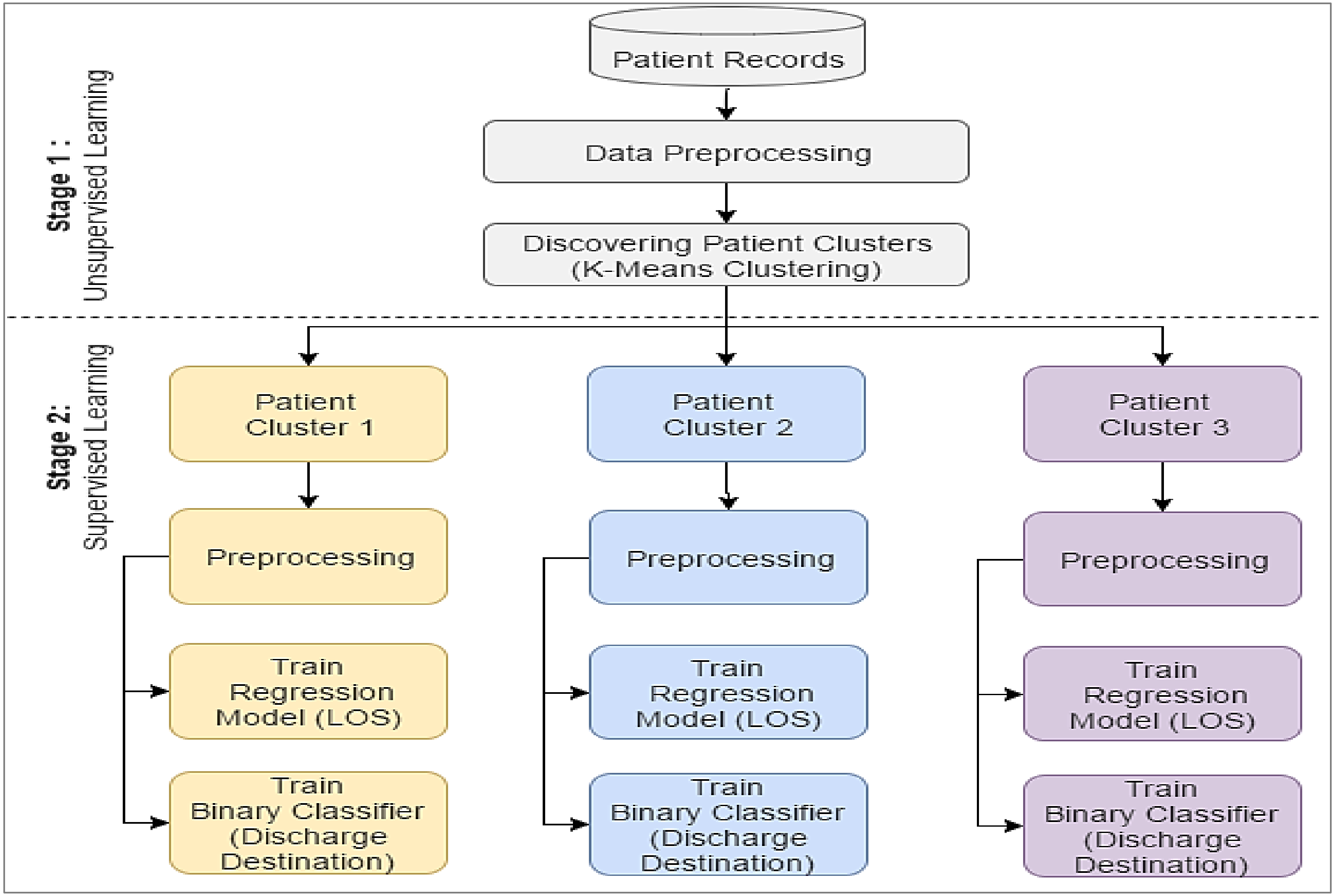
- A significant measure of patient outcomes [1-3].
- A valid proxy to measure the consumption of hospital resources.
- Reported as the main component of the overall cost of hip fracture care [4].

Discharge Destination:

- Having a strategic importance to estimate the needed capacity of long-stay care facilities such as nursing homes.

OÉ Gaillimh
NUI Galway

Approach Overview

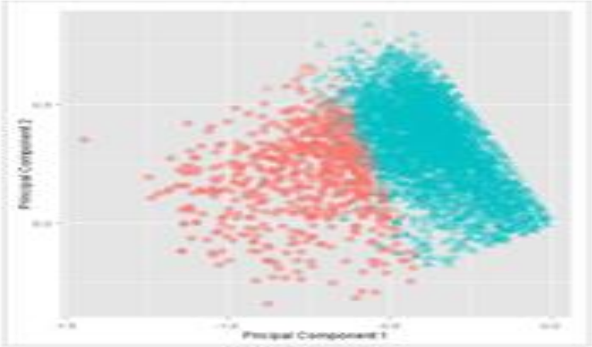


Data Description

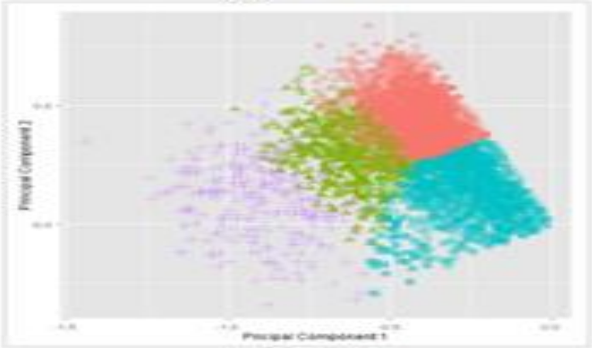
- Irish Hip Fracture Database (IHFD).
- Patient records in the years 2013-2014.
- Patients aged 60 and over.
- 38 data fields such as gender, age, type of fracture, date of admission, and LOS.

OÉ Gaillimh
NUI Galway

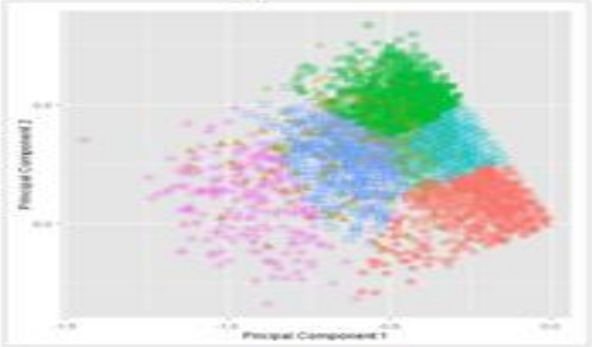
K-Means Clustering Experiments

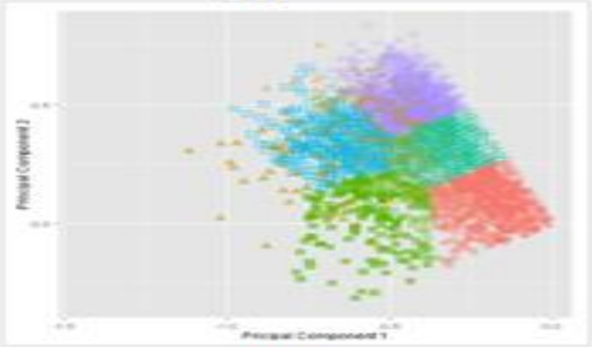

(a) K=2


(b) K=3

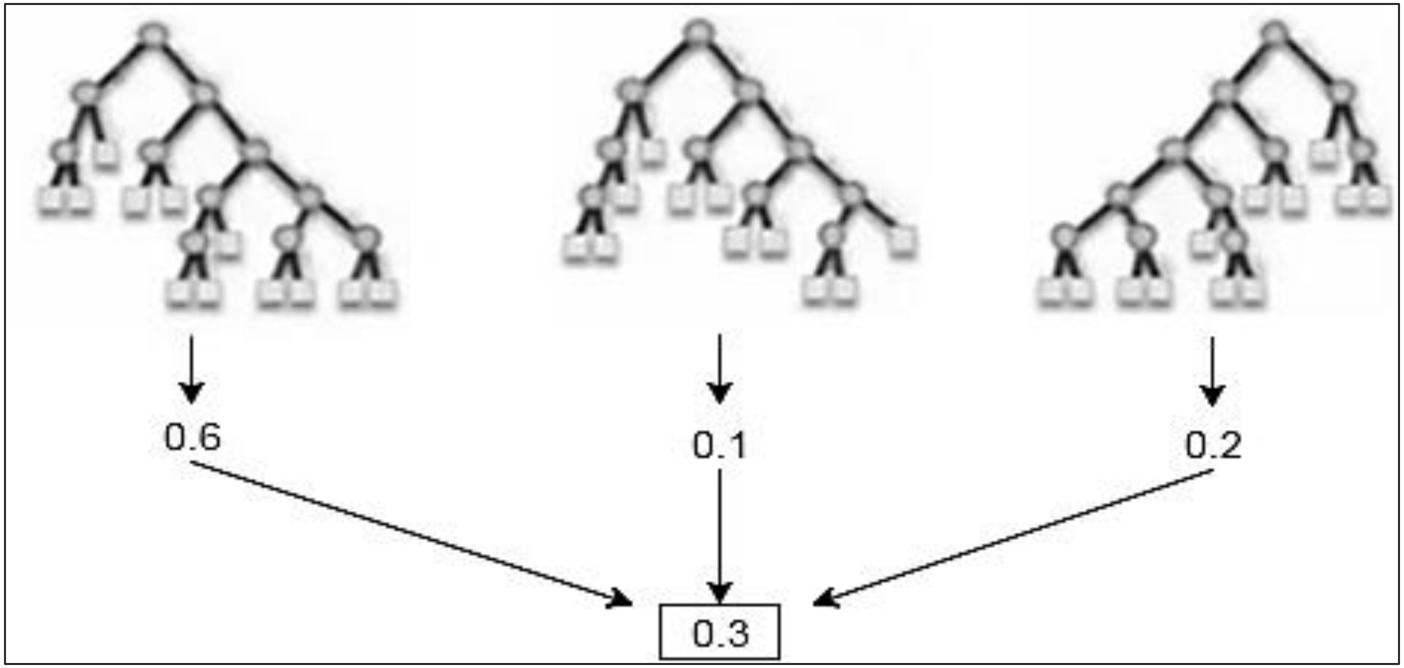

(c) K=4


(d) K=5


(e) K=6


(f) K=7

Learning Algorithm: Random Forests



$$p(c|v) = \frac{1}{T}\sum_{t=1}^{T} p_t(c|v)$$

, where $p_t(c|v)$ denotes the posterior distribution obtained by the t-th tree.

Learning Algorithm: Random Forests

Scholar

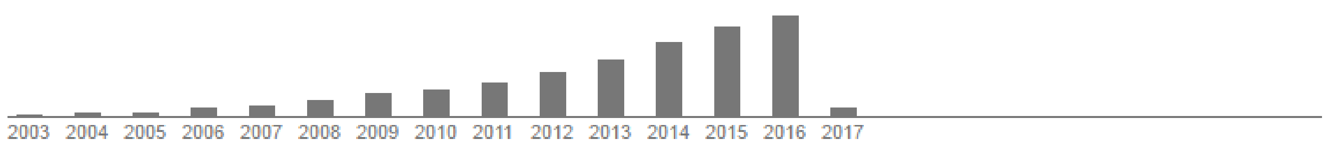
Export

Leo Breiman 1928-2005

Random forests

[PDF] from univ-toulouse.fr

Authors	Leo Breiman
Publication date	2001/10/1
Journal	Machine learning
Volume	45
Issue	1
Pages	5-32
Publisher	Springer Netherlands
Description	Abstract Random forests are a combination of tree predictors such that each tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest. The generalization error for forests converges as to a limit as the number of trees in the forest becomes large. The generalization error of a forest of tree classifiers depends on the strength of the individual trees in the forest and the correlation between them. Using a random selection of features to split each node yields error rates ...
Total citations	Cited by 26024



Source: https://scholar.google.com/citations?user=mXSv_1UAAAAj

Paying Tribute to Leo Breiman (1928-2005)

Leo Breiman 1928-2005

Follow

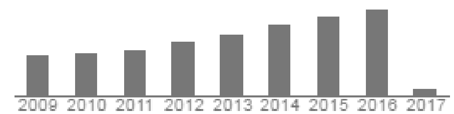
Professor of Statistics, UC Berkeley
Data Analysis, Statistics, Machine Learning
Verified email at stat.berkeley.edu - Homepage

Title 1–20	Cited by	Year
Classification and Regression Trees L Breiman, JH Friedman, RA Olshen, CJ Stone CRC Press, New York	~~31499~~ *	1999
Classification and regression trees L Breiman Chapman & Hall/CRC	31499	1984
Random forests L Breiman Machine learning 45 (1), 5-32	26024	2001
Bagging predictors L Breiman Machine learning 24 (2), 123-140	15844	1996
Estimating optimal transformations for multiple regression and correlation L Breiman, JH Friedman Journal of the American Statistical Association, 580-598	1743	1985
Statistical Modeling: The Two Cutures L Breiman	1566 *	2003
Statistical modeling: The two cultures (with comments and a rejoinder by the author) L Breiman Statistical Science 16 (3), 199-231	~~1562~~	2001
Prediction games and arcing classifiers L Breiman	1480 *	1997

Google Scholar

Get my own profile

Citation indices	All	Since 2012
Citations	90261	43163
h-index	49	32
i10-index	78	47



Feature Selection & Extraction

LOS Regression Model:

- Age
- Patient Gender
- Fracture Type
- Hospital Admitted To
- ICD-10 Diagnosis
- Fragility History
- Time to Surgery

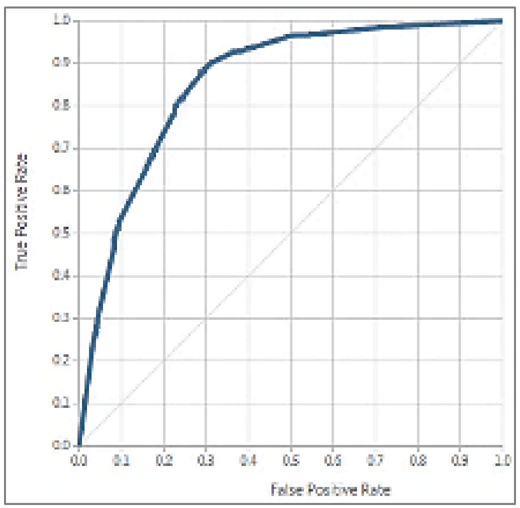
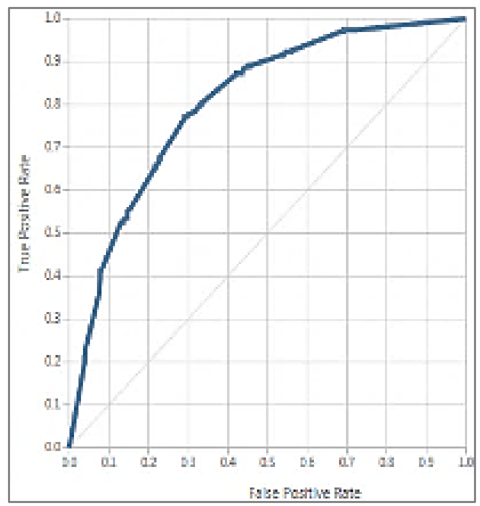
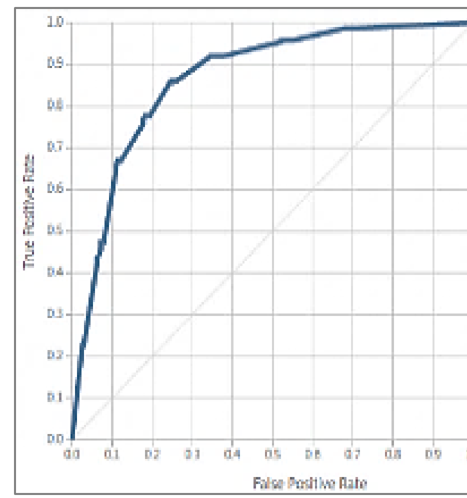
Discharge Destination Classifier:

- Age
- Patient Gender
- Fracture Type
- Hospital Admitted To
- ICD-10 Diagnosis
- LOS
- Fragility History
- Time to Surgery

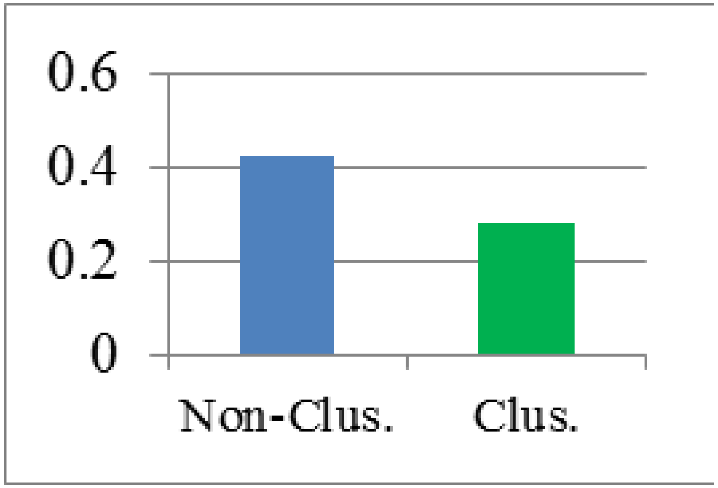
Experimental Results: Regression Accuracy (10-Fold Cross-Validation)

LOS Predictors	Relative Absolute Error (≈)	Relative Squared Error (≈)	Coefficient of Determination (≈)
Cluster 1	0.24	0.14	0.86
Cluster 2	0.30	0.19	0.81
Cluster 3	0.28	0.18	0.82

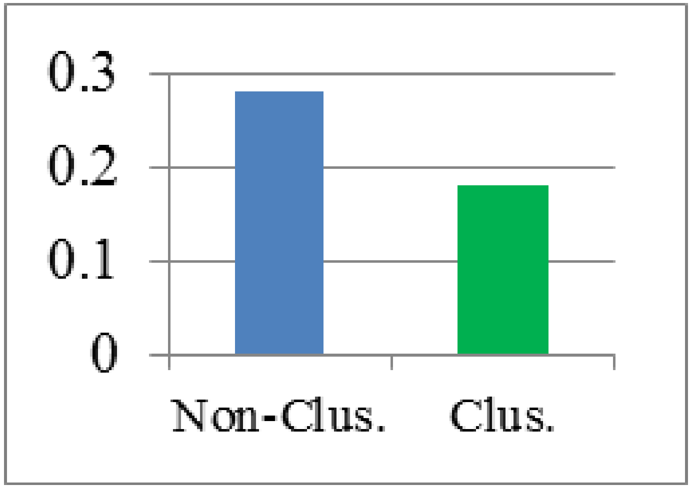
Experimental Results: Classifier Accuracy (10-Fold Cross-Validation)

Cluster1	Cluster2	Cluster3
		
AUC=0.851	AUC=0.739	AUC=0.863

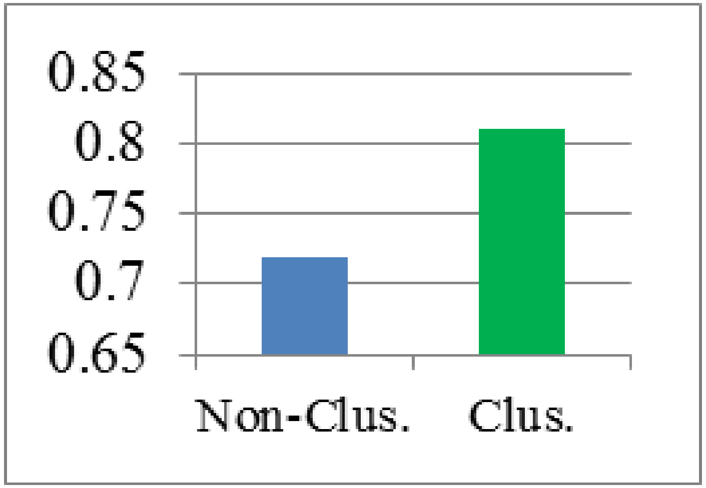
The Significance of Clustering



(a) Relative Absolute Error

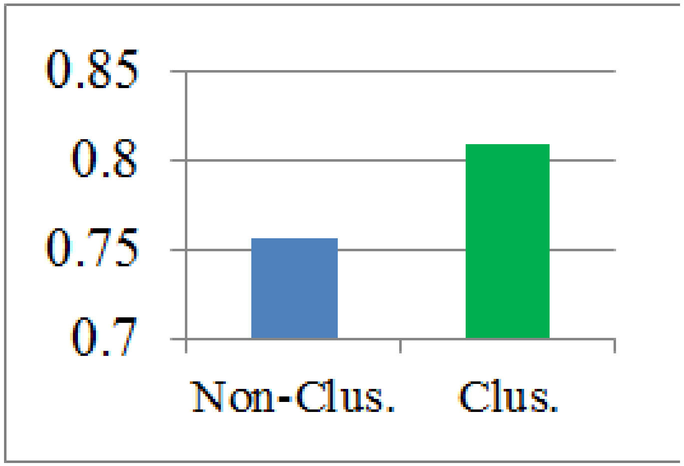


(b) Relative Squared Error

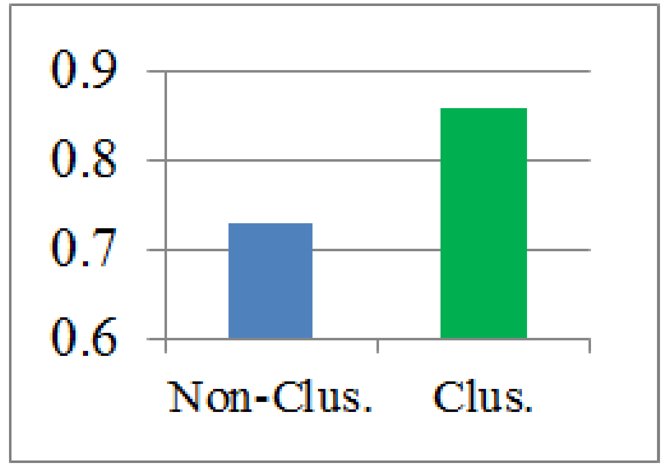


(c) Coefficient of Determination

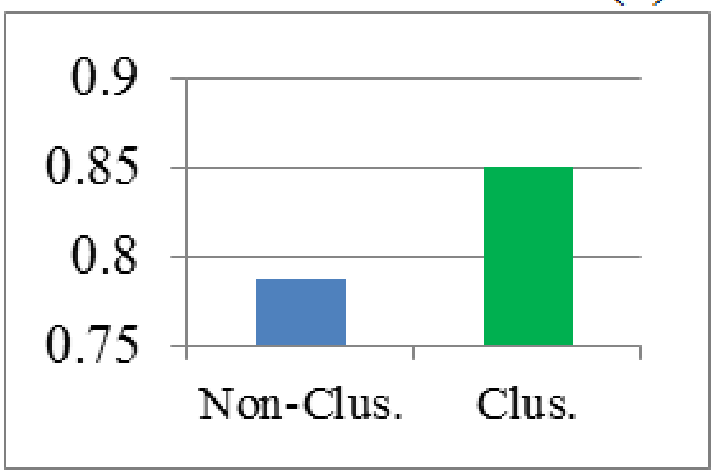
The Significance of Clustering (cont'd)



(a) Precision



(b) Recall



(c) Accuracy

The Significance of Clustering (cont'd)

Feature	Feature Importance Score (≈)		
	Cluster1	Cluster2	Cluster3
Age	0.84	0.49	0.61
Patient Gender	0.14	0.23	0.18
Fracture Type	0.38	0.44	0.21
Hospital Admitted To	0.78	0.93	0.56
ICD-10 Diagnosis	0.48	0.52	0.29
Fragility History	0.44	0.10	0.09
Time To Surgery	0.15	0.27	0.64

* The features were decided based on the permutation importance method [5]

More about Patient Clustering…

ACM DL DIGITAL LIBRARY

Data-driven patient segmentation using K-means clustering: the case of hip fracture care in Ireland

Full Text: PDF Get this Article

Authors: Mahmoud Elbattah National University of Ireland Galway
Owen Molloy National University of Ireland Galway

2017 Article

Published in:
· Proceeding
ACSW '17 Proceedings of the Australasian Computer Science Week Multiconference
Article No. 60

Geelong, Australia — January 30 - February 03, 2017
ACM New York, NY, USA ©2017
table of contents ISBN: 978-1-4503-4768-6 doi>10.1145/3014812.3014874

https://dl.acm.org/citation.cfm?id=3014874

OÉ Gaillimh
NUI Galway

Closing Thought

- *"The key challenges facing healthcare providers in future years are perhaps more organisational and logistical than medical and scientific advances".*

(Sally Brailsford & Jan Vissers 2011)

Source :Brailsford, S. and Vissers, J., 2011. OR in healthcare: A European perspective. *European journal of operational research*, *212*(2), pp.223-234.

OÉ Gaillimh
NUI Galway

References

1. O'Keefe, G.E., Jurkovich, G.J. and Maier, R.V., 1999. Defining excess resource utilization and identifying associated factors for trauma victims. Journal of Trauma and Acute Care Surgery, 46(3), pp.473-478.
2. Englert, J., Davis, K.M. and Koch, K.E., 2001. Using clinical practice analysis to improve care. Joint Commission Journal on Quality and Patient Safety, 27(6), pp.291-301.
3. Guru, V., Anderson, G.M., Fremes, S.E., O'Connor, G.T., Grover, F.L., Tu, J.V. and Consensus, C.C.S.Q.I., 2005. The identification and development of Canadian coronary artery bypass graft surgery quality indicators. The Journal of thoracic and cardiovascular surgery, 130(5), pp.1257-e1.
4. Johansen, A., Wakeman, R., Boulton, C., Plant, F., Roberts, J. and Williams, A., 2013. National Hip Fracture Database: National Report 2013. Clinical Effectiveness and Evaluation Unit at the Royal College of Physicians.
5. Altmann, A., Toloşi, L., Sander, O. and Lengauer, T., 2010. Permutation importance: a corrected feature importance measure. Bioinformatics, 26(10), pp.1340-1347.
6. Breiman, L., 2001. Random forests. *Machine learning*, *45*(1), pp.5-32.
7. Brailsford, S. and Vissers, J., 2011. OR in healthcare: A European perspective. *European journal of operational research*, *212*(2), pp.223-234.

THANK YOU!

mahmoud.elbattah@nuigalway.ie