Exploring antimicrobial resistance prediction using post-hoc interpretable methods

B. Cánovas-Segura¹, A. Morales¹, A. López Martínez-Carrasco¹, M. Campos¹, J. M. Juarez¹, L. López Rodríguez², F. Palacios¹

¹ AIKE research group, University of Murcia

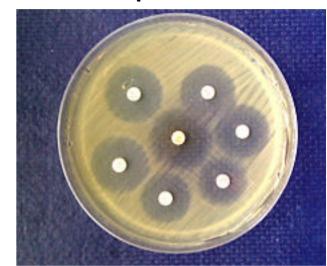
² Hospital University of Getafe

UNIVERSIDAD DE MURCIA



MEDICAL PROBLEM

- Reduce spread of resistant bacteria in hospitals
- AST: Antimicrobial suscep. test
 - Effect of AB on bacteria strains
 - Days to be completed



- Predict if infected by resistant bacteria before AST
- Focus on VRE: enterococcus species resistant to vancomycin

OUTLINE:

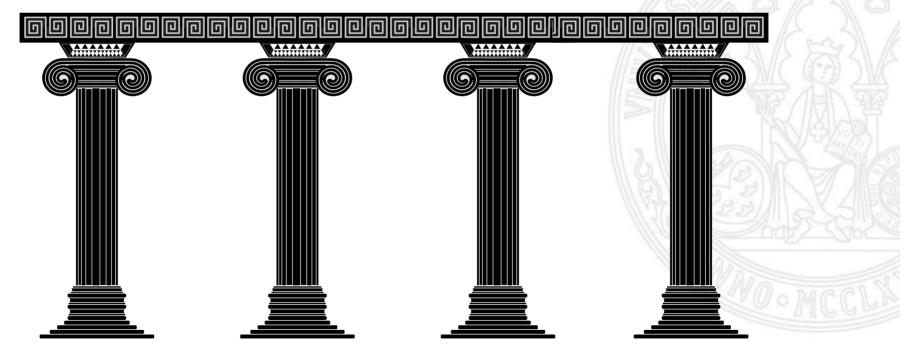
- 1. Medical problem
- 2. Proposal
- 3. Requirements & experiments
- 4. Conclusions & future work



PROPOSAL

- Assess the impact of ML models
- For VRE predictor in a CDSS

User & Data REQUIREMENTS



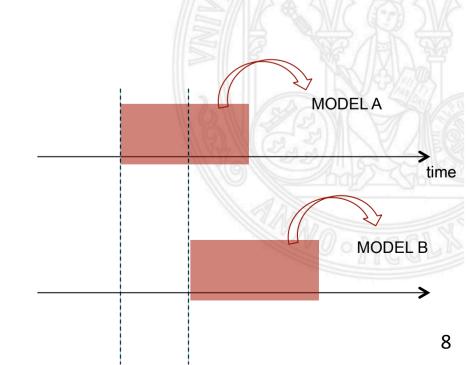
Interpretability Concept drifting High dim. Imbalance classes

- Requirements: INTREPRETABLE
 - Vague term
 - Lipton's classification:
 - Transparent
 - Post-Hoc explanation
 - Trust and easy evaluation
 - Approach:
 - Logistic Regression
 - Decision Trees
 - Rule-based



- Requirements: Concept Drift
 - Influence of factors vary over time
 - Bacteria develop resistance
 - Seasonal variations dissemination bacteria
 - Change clinical protocols

- Approach:
 - Sliding windows
 - Understand & Implement



- Requirements: High Dimensionality
 - Patient clinical record + Lab test results
 - Approach:
 - Filters (FCBF)
- Requirements: Imbalanced Dataset
 - Few observations Resistant class
 - Many Susceptible class
 - Approach:
 - Random oversampling (1:1)



Principal researcher

Other principal researchers



Bernardo

Cánovas-Segura

Place Friday, "I Ristension United Intelligence in January 1991."

Univ. Murcia

The Condensed Nearest Neighbor Rule

The purpose of this note is to introduce the condensed negrest neighbor decision rule (CNN rule) and to pose some unsolved theoretical questions which it mises. The CNN rule, one of a class of mt los decision rules which have appeared in the literature in the past lew years, was motivated by statistical considerations

Manuscript received November 34, 1986; revised October 5, 1867.

newest neighbor decision rale (NN rule). We MN rule and then describe the CNN rule.

The NN rulest C assigns an unclassified sample to the same the respect of a stored, correctly classified supplies. In other collection of a reference points, each classified by arce, a new point is assigned to the same class as but. The most interesting theoretical property of tunder very mild regularity assumptions on the ies, for any metric, and for a variety of less func-imple risk incurred is less than twice the Bayes go of the underlying statistics.) From a practical over, the NN rule is not a prime candidate for properted san rule which receins the basic approprial.

out imposing such stringent atomage requirements. in the CNN rule we first define the notion of a of a sounde set. This is a subset which, when used set for the NN rule, correctly classifies all of the the sample set. A minimal consistent subset is a with a minimum number of elements. Every set oughy, every finite set has a minimal massistant the minimum size is not, in general, schieved N rule uses the following algorithm to determine of the original sample set. In general, however, not find a rainimal consistent subset. We assume caple set is arranged in some order; then we set alled shour and granuag and proceed as follows.

The first sample is placed in grown.

anaple is classified by the NN rule, using as a surrent contents of stores. (Since syons has only suffication is trivial at this slage.) If the second

inductively, the ith sample is obssified by the of stone. If classified correctly it is placed in otherwise it is placed in store.

sthrough the original sample set, the procedure

max (in which case, the consistent subset found to pass is made through GRABHAR with a STORE. (If this happens, all subsequent passes

decision surface has not been changed.)

e all samples had integer-value

o points in the random ordering.

ore realistic experiment was performed using data supplied

ng set. The algorithm terminated after four iterations through

AG, at which time STORE contained 197 of the original 6295

les. An error rate of 1.28 percent was obtained on the indepen-

test set. This was somewhat disappointing in view of the fact a number of simpler classifiers (the ternary reference classi-

linear machine,^[4] and piecewise-linear machine^[4]), using erably less computer time, achieved error rates on the order

5) The final contents of stume are used as reference points for ontents of quarrage are discarded.

the rule behaves as follows: If the Bayes risk is indurlying densities of the various classes have the algorithm will tend to pick out points near y) boundary between the classes. Typically ided within a class will not be transferred to will be correctly classified. If the Bayes risk is Ill contain essentially all the points in the original important reduction in sample size will have theoretical properties of the CNN rule baye been

The CNN rule has been tried on a number of problems, both real and artificial. In order to investigate the behavior of the rule when the classes are (essentially) disjoint—the case in which the CNN rule is of greatest interest-several experiments similar to the following wore run. The underlying probability structure for a two-class problem was assumed to consist of two probability densities, each a uniform distribution on the supports shown in Fig. I. The set of all vectors with integer components lying within each

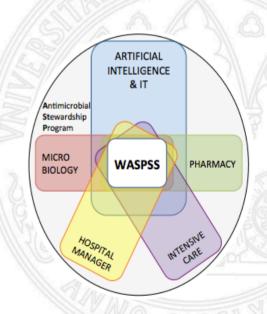
Let X and Y take on the values $\{x_i: i=1, \cdots, m\}$ and $\{y_i: i=1, \cdots, m\}$, respectively, where $n \geq m$. A decision rule for X in terms of Y can be considered as a partition $\{A_i, i=1, \cdots, m\}$ such that $A_i \cap A_j = \phi, i \neq j,$ and $\bigcup_{i=1}^n A_i = \{y_i, j=1, \cdots, n\}$ where the decision is x_i if $Y \in A_i$. This also defines a "post-decision" random ts for the fact that occasionally the decision surface lies y within one or the other of the supports rather than between variable Z, where Z is defined by $Z = z_i$ if $Y \in A_i$, $i = 1, \dots, m$. The points most deeply imbedded within each class were the

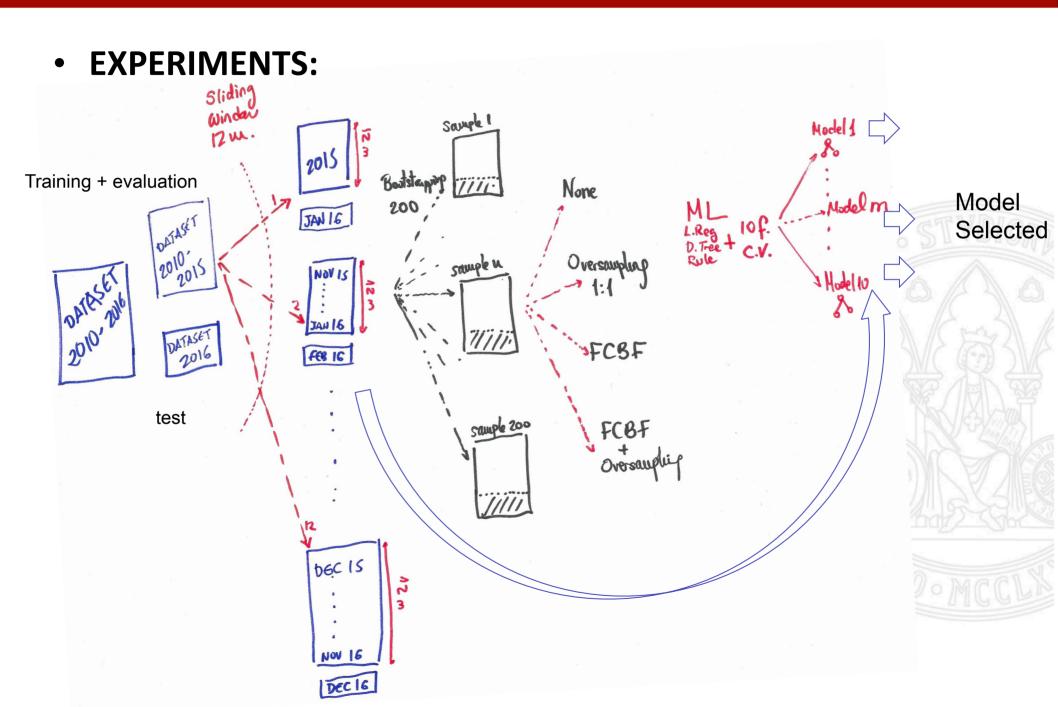
Two putative measures of the efficiency of this system are un tainty (or equivocation) and probability of error. It is desirable to determine the relationship between these two measures. In agy of IBM.^[5] This data consisted of approximately 12 000 nensional binary vectors drawn from 25 different statistical particular, we can compare H(X|Y) with the minimum probability of error $P_0(e)$ if we want to evaluate the channel independent of ations. (The data represent upper-case typewritten characters, ding "I," typed with nine different styles of fonts.) The the decision rule. Otherwise we can compare, given a particular decision rule, H(X|Z) with the probability of error P(e). The pur-) samples were divided into a training set and a testing set of ximately equal size, and the CNN algorithm was used on the pose of the paper is to demonstrate the exact relationship bet H(X|Y) and $P_{\theta}(e)$.

First, we relate $H(X|y_k)$ to $P_0(e|y_k)$ for each k. Now $P_0(e|y_k) = 1 - \max_i P(x_i|y_k)$, and letting y_k be fixed, we denote $P_i = P(x_i|y_k)$ $i=1,\cdots,m$, such that $P_1\geq P_i,\,i=2,\cdots,m$. Then $P_0(e|y_k)=$

EXPERIMENTS

- VRE:
 - Bacteria: Enterococci faecium + Enterococci faecalis
 - Antibiotic: Vancomycin
- Dataset:
 - WASPSS platform
 - University Hospital Getafe
 - 2010-2016
 - 1393 + Enterococci cultures (42 VRE)
 - 571 attributes + binary class (resistant|suscept.)







Mean (95% C.I) number of predictors per model

None

75 -

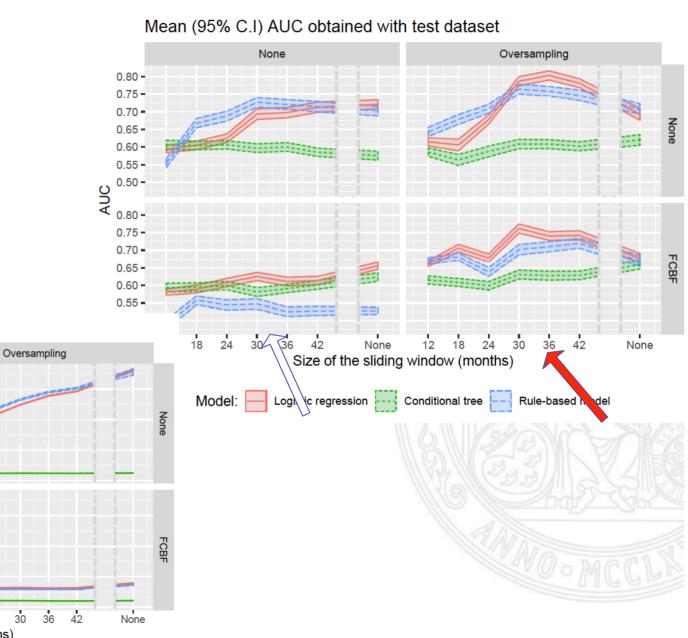
50 -

Number of predictors

25 -

0 -

18



Size of the sliding window (months)

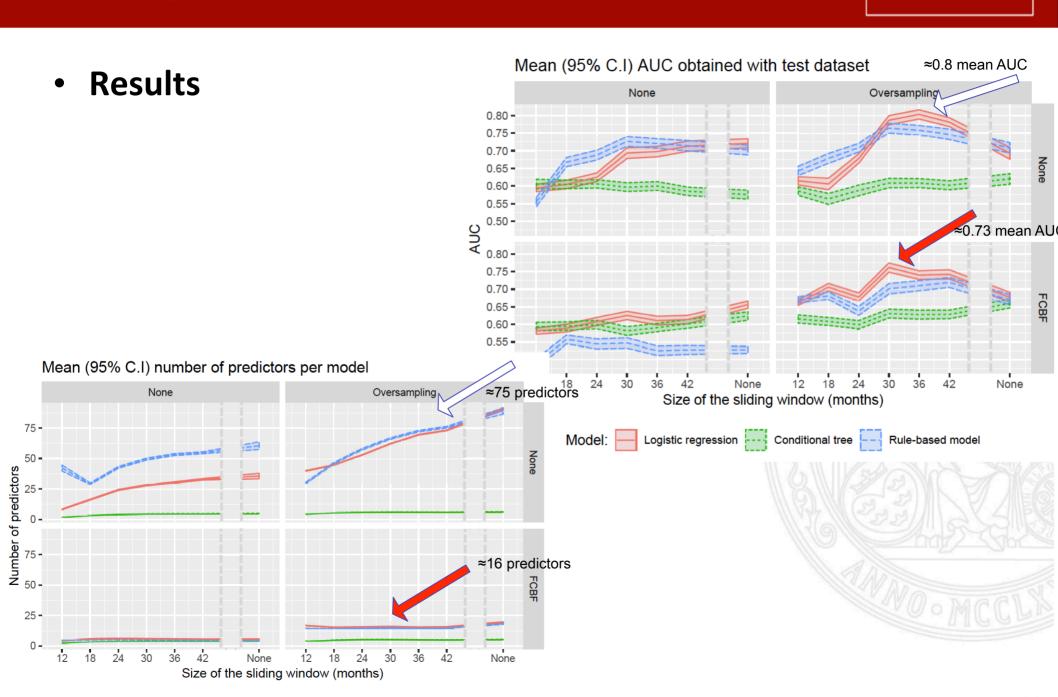
36

None

Logistic regression ---- Conditional tree

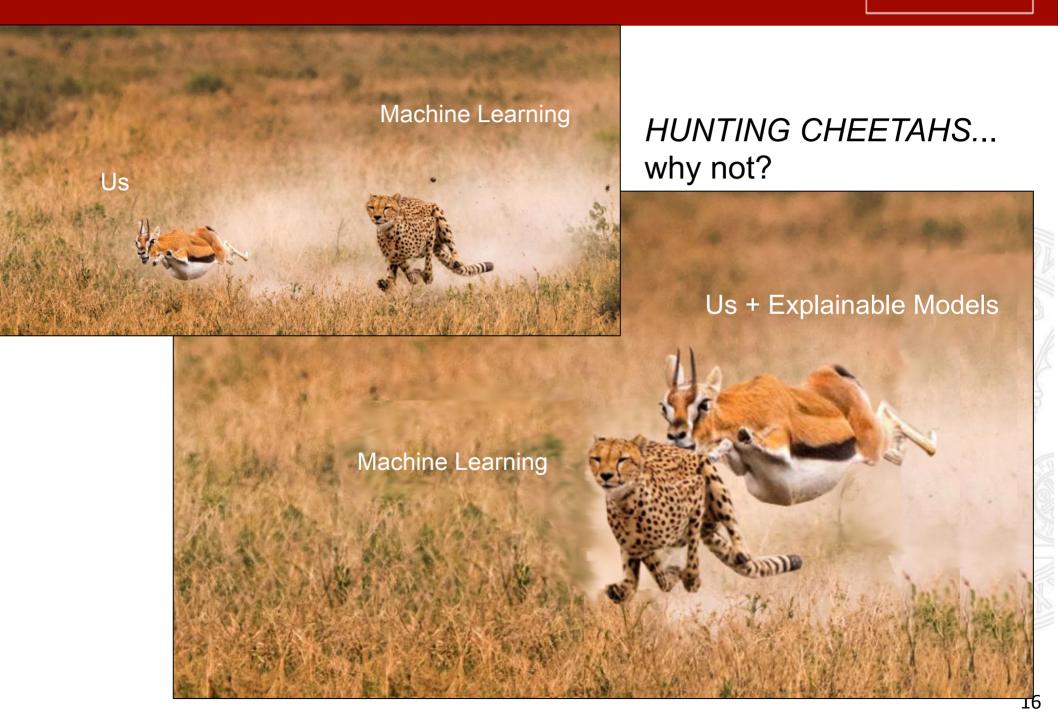
Rule-based model

Model:



Conclusions & future work

- 3 interpretable models for VRE problem
- Temporal dependence: concept drifting
- ✓ Oversampling +FCBF in VRE
- Performance vs. interpretability (80%+75p. vs. 73%+16p.)
- How evaluate interpretability of models
- Models in CDSS and clinical activity flow





Dziękuję za uwagę

(THANK YOU FOR YOUR ATTENTION)

Jose M. Juarez

jmjuarez@um.es