DeathPredict Report

Name:	Pierre-alexandre BARIL, Morgan VATERKOWSKI, Quentin LEFORT
Research Centre:	EPITA
Research Project Title:	DeathPredict
Primary Supervisor:	Antoine Liutkus

1 Abstract

The predictors of in-hospital mortality for intensive care units (ICU)-admitted HF (Heart Failure) patients remain poorly characterized. This is why we aim to predict if a patient will die from heart failure(HF) while they are in intensive care unit (ICU). For that we use the patient's different constant like 'age', 'diabetes', 'calcium in blood' and many others (48 of them).

2 Introduction

Introduction, Système, Implémentation, Experiences, Discussion+Conclusion, Références We aim to predict if a patient will die from heart failure(HF) while they are in intensive care unit (ICU). For that we use the patient's different constant like 'age', 'diabetes', 'calcium in blood' and many others (48 of them).

The repository is on github[1] and the code was developped with the help of kaggle[2].

3 Data

3.1 Data Source and collection

The MIMIC-III database (version 1.4, 2016) is a publicly available critical care database containing de-identified data on 46,520 patients and 58,976 admissions to the ICU of the Beth Israel Deaconess Medical Center, Boston, USA, between 1 June, 2001 and 31 October, 2012. These data include comprehensive information, such as demographics, admitting notes, International Classification of Diseases-9th revision (ICD-9) diagnoses, laboratory tests, medications, procedures, fluid balance, discharge summaries, vital sign measurements undertaken at the bedside, caregivers notes, radiology reports, and survival data12. After successful completion of the National Institutes of Health Protecting Human Research Participants web-based training course, we obtained approval to extract data from MIMIC-III for research purposes (Certification Number: 28860101).

3.2 Final Version

The data we used is a kaggle dataset[3] . The data come under an Excel (csv) file containing 51 columns for each of the 1177 patients.

4 Implementation

We decided to code under the Python language. Moreover, we used the machine learning package : sci-kit learn.

5 Experiences

5.1 MLP

The first idea is to use a Multi-Layer Perceptron. However, this implementation gives poor results : 40 percent accuracy. Antoine advised to use trees, therefore leadind to the next experience.

5.2 Trees

1. Decision Tree
2. Random Forest

5.3 All models

We used Pycaret module to train all models on our data in order to compare them all. It can be found here[4].

6 Discussion & Conclusion

Interstingly enough, the naive bayes model is the best for predicting death and survival, which is equivalent for the F1-score.

Notes

1. See https://github.com/PierreAlexandreDev/DeathPredict/ for more information.

2. See https://www.kaggle.com/pierrealexandre78/deathpredict for more information about the main code and history.

3. See https://www.kaggle.com/saurabhshahane/in-hospital-mortality-prediction

4. Pycaret Module: https://www.kaggle.com/sonalisingh1411/pycaret-automl-heart-failure-prediction