INTRODUCTION
Data mining is not new and has been efficiently deployed in other domains (Bank, Cybernetics, Marketing, Energy, etc.). The amount of data collected in medical information systems is tremendous, but much data remains unused because of its complexity. This still increasing volume of data requires new analytical approaches that are efficient, sensitive and better than classical statistical techniques to handle Big Data.

METHOD DIVERSITIES BETWEEN DATA MINING AND STATISTICS

OBJECTIVE
The main objective of this study is to assess the capability of state-of-the-art data mining techniques to be applied on a healthcare database in order to better understand drivers of healthcare expenditure and the management of diseases. We utilized the approach in order to explain the patient features that are the main drivers of cost associated with HIV patients’ hospital management. A secondary objective is to assess how reproducible such an approach is on other medical databases with potentially different patient features.

METHODS
KEY FIGURES

DATA EXTRACTION
The PMG-ASD (Unveil Medical Information System – Medicine, Surgery, Oncology) (until) database was used to extract all hospital stays in 2013 with at least one of the following HIV ICD-10 (International Classification of Diseases, 10° revision) codes as principal diagnosis related to diseases or significantly associated diagnosis (V12, Y02.8, Z02.6, Y03, B08, B01).

RESULTS


DISCOVERY OF PATIENT PROFILES
1 - Infections contributed to HIV? 2 - Bacteri or opportunistic infection? 3 - Presence of anemia? 4 - Fungal opportunistic infection? 5 - Inverse non-metabolic comorbidity?

DATA MINING ALGORITHM
Data were analyzed with a supervised-learning data mining technique, an enhanced Decision Tree algorithm based on breast CAIT-model (Figure 3).

1 YEAR FOLLOW-UP OF PATIENTS & CLASSIFICATIONS OF PATIENTS
For each patient hospitalized once in 2013 with an HIV code, a 12-month follow-up was performed to capture any further hospital stay. The same algorithm and medical review as for the inclusion stay were conducted on these additional stays, in order to evaluate stays not related to HIV. All stays were classified by HIV clinician experts into 10 groups of comorbidities and 5 groups of opportunistic infections (OIs) (Figure 2).

HOSPITALIZATION COST
The cost of stays were evaluated from a NHS perspective. The PMG database provides many details about stays’ cost as it was specifically designed for economic evaluation of hospital activities.

OUR DATA MINING ALGORITHM

DATA SCIENTIST

ADVANCED DATA MINING APPROACH
Prediction of costs associated with the hospital management of HIV patients in France

CONCLUSION
The present study is a pilot that successfully demonstrates how a data mining technique, from the field of Artificial Intelligence, can help us better understand hospital costs for HIV patients: our data mining algorithm identified specific patient profiles which explain the differentiating cost drivers in HIV inpatient care. Classical statistical approaches would struggle to provide such detailed profiles with numerous combinations of variables.

1 - This approach could work with other data sources, especially with more clinical and laboratory data (e.g. virology).
2 - It can be adapted to drive any available target variable (e.g., doctors, patients’ unrecorded suppressed or patients dropping out of care).
3 - Such collaborative efforts between healthcare professionals and engineers can lead to enhanced uses of health data that can provide new answers and thus improve disease management.
4 - Preventing a opportunistic infection is really diagnosed and patients with diabetes-Metabolic’s comorbidity, is much more cost reducing than preventing a fungal infection.
5 - Based on results of Data Mining analyses, clinicians will know the medical interventions to focus on and prioritize.

REFERENCES:
1. Adem Karahoca, Dilek Karahoca and Mert Şanver (2012). Survey of Data Mining and Applications (Review from ... Data Mining Applications in Engineering and Medicine, Associate Prof. Adem Karahoca (Ed.), InTech, DOI: 10.5772/48803.

DATA MINING PARADIGM

DATA SCIENTIST

ADVANCED DATA MINING APPROACH
Prediction of costs associated with the hospital management of HIV patients in France

CONCLUSION
The present study is a pilot that successfully demonstrates how a data mining technique, from the field of Artificial Intelligence, can help us better understand hospital costs for HIV patients: our data mining algorithm identified specific patient profiles which explain the differentiating cost drivers in HIV inpatient care. Classical statistical approaches would struggle to provide such detailed profiles with numerous combinations of variables.