Challenges to Turn the Bosch Dataset into Time Series
Sid Ahmed MOKEDDEM, Mouhamadou Saliou DIALLO, Agnès BRAUD, Gabriel FREY and Nicolas LACHICHE
University of Strasbourg
{mokeddem, ms.diallo, agnes.braud, g.frey, nicolas.lachiche}@unistra.fr

ABSTRACT
Factories are interested to maximize their production and availability of their production machines. The production lines data are recorded for every production process as they progress through each station. The data provides information about product quality failure. Most of existing studies presents the prediction when the production fails the quality control which is not beneficial from production point of view. Our aim is to change the Bosch data set representation to obtain sequential data representations to build up more relevant prediction model. This paper describes the challenges of changing data representation of the Bosch production line performance.

DATA SET

*Bosch monitoring system produced a huge data of its manufacturing process and shared them through Kaggle challenge
*Bosch aims through this challenge to have an efficient prediction model of their production quality.
*Most of existing studies developed an instant prediction model to find the failed products once its production process is finished which is not more relevant.
*It is more interesting to build models to predict in advance the production failure by analyzing the stations health status in order to prevent their failures by using TS prensentation.

CHALLENGE

*Stations have different frequencies, meaning that pieces follow different flows, not on all stations.
*A total of 7148 flows are identified which means that there are several product families in the data set.
*By clustering the products into product families and be used to achieve the best prediction performance.
*The feature vectors having a huge dimension and very little number of non-zero values.
*Timestamp features are used to identify the time when the feature value of station was collected.
*A sequential representation, more precisely a Time Series representation, to predict the remaining useful life for each station.

ACKNOWLEDGEMENT
This work has received funding from INTERREG Upper Rhine (European Regional Development Fund) and the Ministries for Research of Baden-Württemberg,Rheinland-Pfalz (Germany) and from the Grand Est French Region in the framework of the Science Offensive Upper Rhine HALFBACK project.