1. Filling missing data:  
   <https://www.nature.com/articles/s41592-021-01302-4>  
   The class imbalance problem  
   synthetic minority over-sampling technique (SMOTE)  
   <https://ying-ju.github.io/subsampling.github.io/>.
2. Feature Reduction:  
   <https://scikit-learn.org/stable/modules/feature_selection.html>  
   Scikit-learn
3. To evaluate the impact of this new pre processing under the impact of XGBoost (reported: [file:///C:/Users/pedra/Downloads/TSP\_CMC\_43880.pdf](about:blank)).
4. We believe Random forest and Naïve Bayes could perform better (why? TODO) (<https://scikit-learn.org/stable/modules/naive_bayes.html>)

* XGBoost in SCikit-learn (https://towardsdatascience.com/getting-started-with-xgboost-in-scikit-learn-f69f5f470a97)