**Meeting Data Mining Group 23.12.2018**

**Luogo**: Online, Skype

**Data**: 23.12.2018

**Presenti**: Daniele Gadler, Riccardo Manetti, Maddalena Amendola

**Assenti giustificati**: Gemma Martini

Punti da discutere:

1. Aggiunta account di ognuno al Team su Kaggle. Possibilmente condividere account per fare più submission in un giorno
2. **Clustering**:
   1. Discussione della trasformazione del “ps” in un valore compreso tra 0 e 1. Ripetere la creazione dei cluster con tale attributo.
   2. Creazione di nuovi attributi? (EX: In base allo status, aggiungere una nuova colonna con il numero di membri per customer)
3. **Classification:** 
   1. Decision Tree: Okay! Bisogna aggiungere una descrizione del decision tree ottenuto e delle metriche ottenute.
   2. Discussione sul fatto se vogliamo trovare customer che NON vanno in default (0 as default) o customer che VANNO in default (1 as default)
   3. Random Forest: utilizzo di imblearn per fare oversampling / undersampling. Che tecnica utilizzare per ribilanciare il dataset? 10-fold crossvalidation.?
   4. Deep Learning: Fare qualche esperimento, ma non perderci troppo tempo con dsklearn e multi-layer perceptron
4. **Association Rules Mining**
   1. Frequent patterns extraction with different values of support and different types (i.e. frequent, close, maximal) and discussion of the most interesting results obtained. Ù
   2. Association rules extraction with different values of confidence (6 points)