

FINAL CASE STUDY : PREDICTIVE ANALYTICS AND DATA MINING

Every team selects a dataset (kaggle strongly suggested) and makes a complete analysis with prediction

these are the chapters of the final document (ppt slides or word doc)

the chosen dataset should be not less than 5000 rows and at least 15 columns

1. introduction : What is the topic ? why is it important ? what has to be predicted ?
The data : what kind of data are in the datasets (columns , lines , ...), what is the variable that will be predicted ? which are the features for prediction ?
2. Exploratory analysis : some statistical information about variables (statistics, summary, first graph , ...)
3. Detailed analysis : the purpose of this is to determine influence of features on variable to predict , correlation matrix, graph, aggregation etc...it is strongly suggested to present a screenshot of your flow.
4. Prediction : use the ML Analysis as seen in class with Dataiku try to get the highest score , improve the algorithm perf , show perf metrics and variable importance.
5. Conclusion : Does the detailed analysis and the prediction are converging ?, what recommendations could you make based on your analysis ? (recommendations means : some ideas or proposal you could provide to managers or stakeholders based on your analysis.)

Make all screenshots of dataiku including the flow : graphs , corr matrix , models performance , confusion matrix

Variable importance , subpopulation analysis , etc...

If you can produce a map and display on it your datas your grade will be augmented same if you choose a subject allowing to join several datasets