

|  |
| --- |
| Auteur1  [auteur1@](mailto:auteur1@i)imt-atlantique.fr  Auteur2  [auteur2@i](mailto:auteur1@telecom-bretagne.eu)mt-atlantique.fr  Auteur3  [auteur3@i](mailto:auteur1@telecom-bretagne.eu)mt-atlantique.fr |

|  |
| --- |
| Titre du document Rapport  Version 0.1 - 01/01/2017  Formation d'ingénieur  Année scolaire 2017-2018 |

Résumé

Ce document est un modèle de rapport compatible avec la charte graphique de IMT Atlantique.

Mots clés : traitement de texte, styles, modèles

Abstract

This document is a report template compliant with the IMT Atlantique corporate identity & style guide.

Keywords : word processor, styles, template

Sommaire

[Résumé 3](#__RefHeading__20344_1162629425)

[Abstract 3](#__RefHeading__20346_1162629425)

[I. Introduction 6](#__RefHeading__20348_1162629425)

[II. Le projet 6](#__RefHeading__20350_1162629425)

[II.1 Aspects techniques 6](#__RefHeading___Toc1855_245993825)

[II.1.1 Le logiciel 6](#__RefHeading___Toc1857_245993825)

[II.1.2 Le hardware 6](#__RefHeading___Toc1859_245993825)

[II.2 Aspects économiques 6](#__RefHeading__20462_1162629425)

[II.2.1 Coûts de production 6](#__RefHeading__20572_1162629425)

[II.2.2 Etude de marché 6](#__RefHeading___Toc1861_245993825)

[III. Conclusion 6](#__RefHeading__20352_1162629425)

[Références Bibliographiques 7](#__RefHeading__20354_1162629425)

[Annexes 9](#__RefHeading__23102_1162629425)

[Annexe 1 10](#__RefHeading__23104_1162629425)

[Annexe 2 10](#__RefHeading__20574_1162629425)

# Introduction

As data scientist at LinkedIn, our colleagues from Marketing want to organize an online marketing campaign for their client, a restaurant in the Bay area. They have read a lot about social network analysis and asked us to find the 5 most influential people on the network who would best promote the restaurant.

When taking a closer look at the dataset we are given, we can see that only 40% of users have provided their location. We will therefore have to fill the uncomplete profiles before finding the influencers.

In order to do that we are given a dataset representing a LinkedIn graph, each node being a person and edges the fact they are connected. People also have three attributes: location (locations they live/have lived in), college (college(s) they went to), employer (company(ies) they worked for). Only 40% of the people in the graph have provided their attribute.

Our aim is to use connections and correlations between those attributes and people in order to fill the missing profiles, and use those results in order to find influencers.

# problem enunciation and statistics.

## problem enunciation and task analysis



As data-scientists, we intend to use the CRISP-DM methodology in order to solve the problem, which consists in different phases as presented in the chart beside.

First of all, we needed to do some researches about similar studies, because it may help us understand more the problem and see different solutions. We used several papers but the most useful was [1] User Profiling in an Ego Network: Co-profiling Attributes and Relationships.

We should now list all the factors we think might lead a user to have a certain attribute, as we will use it later. This will help us understand the statistics that are relevant to draw from the graph.

We then have to understand how the dataset is made so that we can use the data properly.

We then draw the statistics and use the statistics to elaborate one or several models for user profiling.

Once it is done, we have to implement the model using the data given and assess the model by comparing our predictions with the ground truth. Once it is done for every model, we will know which one are the best, meaning the one with best accuracies.

We will now have the filled profiles and will be able to use it to establish who are the influencers that might help us find a restaurant at the bay.

## Statistics

We therefore want to draw interesting statistics from the graph that might help us understand what factors make that a node has a certain attribute.

# Conclusion

Références Bibliographiques

Annexes

En général, on regroupe toutes les annexes en fin du document.

Annexe 1

* + Annexe 2
    1. Annexe 3

|  |
| --- |
|  |