



Automatic data reporting and visualisation

IT project

Houssam Bahhou
Julien Briand

onepoint.

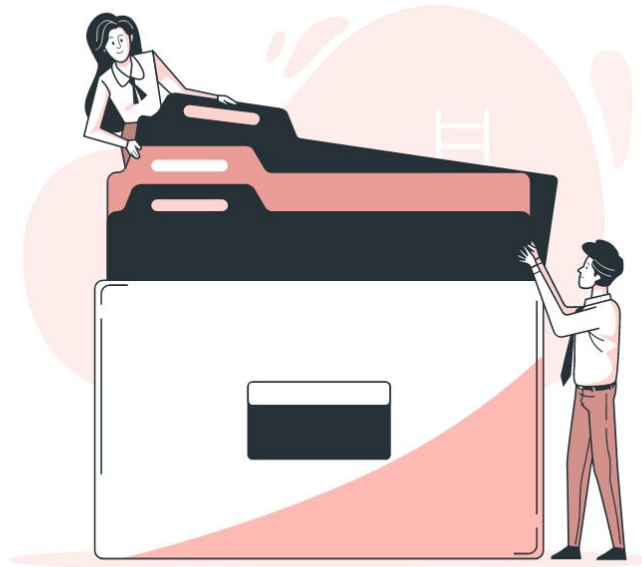
Table of contents

01 Client requirements

Client presentation and objectives

03 Interface presentation

Presentation of the final deliverable



02 Progress

Presentation of the work done and choices made

04 Expected projection

Possible outcomes of the project

01



Client requirements

Client presentation



Company vision & objectives :

Consulting company which helps others to develop in a technical ways through models conception in particular.

Team : Maxime Adolphe

Silvia Ciappelloni

Coralie Vennin

Objectives



Develop a proof of concept :

- An interface to help consultant analyse his data :
 - Import the chosen data and quick graphic return
 - Analyse data in a statistical way
 - Provide a model for an optimal prediction of the features required



Analyse the data given and provide a model to predict the satisfaction level of customers

Database given : survey on public transport

	Horodateur	Quel âge avez-vous ?	De quel département êtes-vous ?	Comment décririez-vous votre usage des transports routiers en NA ?	Quels sont vos usages de transport en commun	À quelle fréquence utilisez-vous les transports routiers de votre région ?	Quelle application utilisez-vous pour réaliser un itinéraire de transport ?	Saviez-vous que vous pouviez réaliser un itinéraire depuis le portail de la Nouvelle Aquitaine ?	Aller sur le site transport de la NA, puis indiquez quelle est votre niveau d'appréciation globale du portail ?	En allant sur le portail de la nouvelle Aquitaine, quelles sont les informations que vous souhaiteriez trouver ?	...	3	4	5	6	7	8	9	10	Quelle est probablement que vous recommanderiez le portail de la Nouvelle Aquitaine à un ami ou collègue ?
0	10/4/2021 12:37:21	37	33	Utilisait TBM pour faire des itinéraires et vo...	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	\n\nN'est pas jamais aller - ne savait même pa...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
1	10/4/2021 18:15:12	24 ans	33	Très peu \n\nLe de temps en temps ex 5 c par s...	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	Horaires \n\nTrajets \n\nTarifs pour les sévic...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
2	10/4/2021 18:28:10	30	33	N'utilise pas - voir le moins possible	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	Horaires\n\nItinéraire \n\nPrix \n\nbouchon/ c...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N
3	10/11/2021 18:39:13	27	33	1 fois par mois	Activités de la vie quotidienne (aller au spor...	NaN	TBM, Google map, Le portail de la Nouvelle Aqu...	NaN	1	Achat de billet en ligne sans devoir faire val...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	N

27 columns &

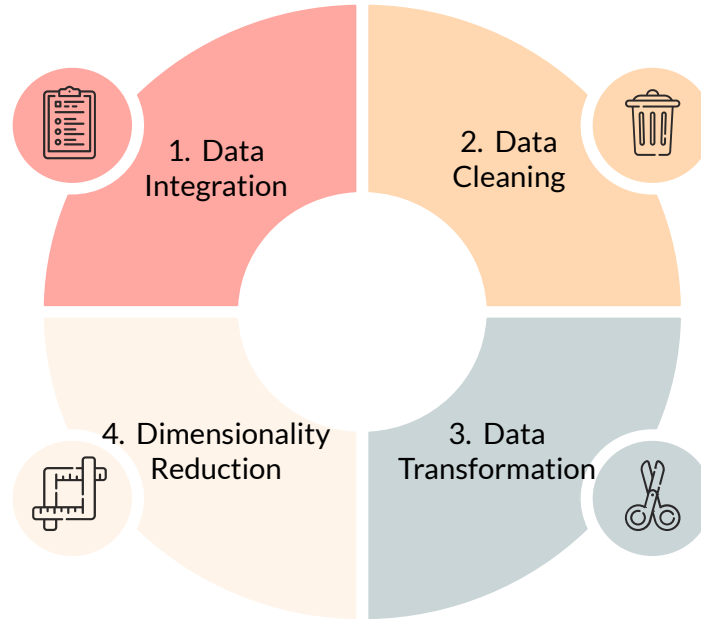
29 answers



02

Progress

Data pre-processing



Data cleaning

Renaming attributes

	Horodateur	Quel âge avez-vous ?	De quel département êtes-vous ?	Comment décrieriez-vous votre usage des transports routiers en NA ?	Quels sont vos usages de transport en commun	À quelle fréquence utilisez-vous les transports routiers de votre région ?	Quelle application utilisez-vous pour réaliser un itinéraire de transport ?	Saviez-vous que vous pouviez réaliser un itinéraire depuis le portail de la Nouvelle Aquitaine ?	Aller sur le site transport de la NA, puis indiquez quelle est votre niveau appréciation globale du portail ?	En allant sur le portail de la nouvelle Aquitaine, quelles sont les informations que vous souhaiteriez trouver ?	...	3	4	5	6	7	8	9	10	Quelle est la probabilité que vous recommandiez le portail de la Nouvelle Aquitaine à un ami ou un collègue?	Souhaitez-vous partager un commentaire, remarque ou suggestion ?
0	10/4/2021 12:37:21	37	33	Utilisait TBM pour faire des itinéraires et vo...	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	\n\nN'est pas jamais aller - ne savait même pa...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
1	10/4/2021 18:15:12	24 ans	33	Très peu \n\nLe de temps en temps ex 5 c par s...	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	Horaires \n\nTrajets \n\nTarifs pour les sévic...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
2	10/4/2021 18:28:10	30	33	N'utilise pas - voir le moins possible	NaN	NaN	Google map, TBM	Non, je ne le savais pas	3	Horaires\n\nitinéraire \n\nPrix \n\nbouchon/ c...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
3	10/11/2021 18:39:13	27	33	Activités de la vie quotidienne (aller au spor...	NaN	NaN	TBM, Google map, Le portail de la Nouvelle Aqu...	NaN	1	Achat de billet en ligne sans devoir faire val...	...	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
4	10/12/2021 10:50:13	23	33	Activités de la vie quotidienne (aller au spor...	Quelques fois par mois	Quelques fois par mois	TBM, Google map, SNCF ou OuiGo	Non, je ne le savais pas	4	Des itinéraires avec les temps les plus courts...	...	5.0	1.0	5.0	2.0	4.0	1.0	5.0	1.0	0	NaN

Filling NaN

Deleting unlabeled

Data transformation

Treating categorical data:

Fréquence		Fréquence
Plusieurs fois par semaine	Mapping →	0
Quelques fois par an		1
Quelques fois par mois		2

Standardizing data:

$$x_{\text{stand}} = \frac{x - \text{mean}(x)}{\text{standard deviation}(x)}$$

Cleaned Data

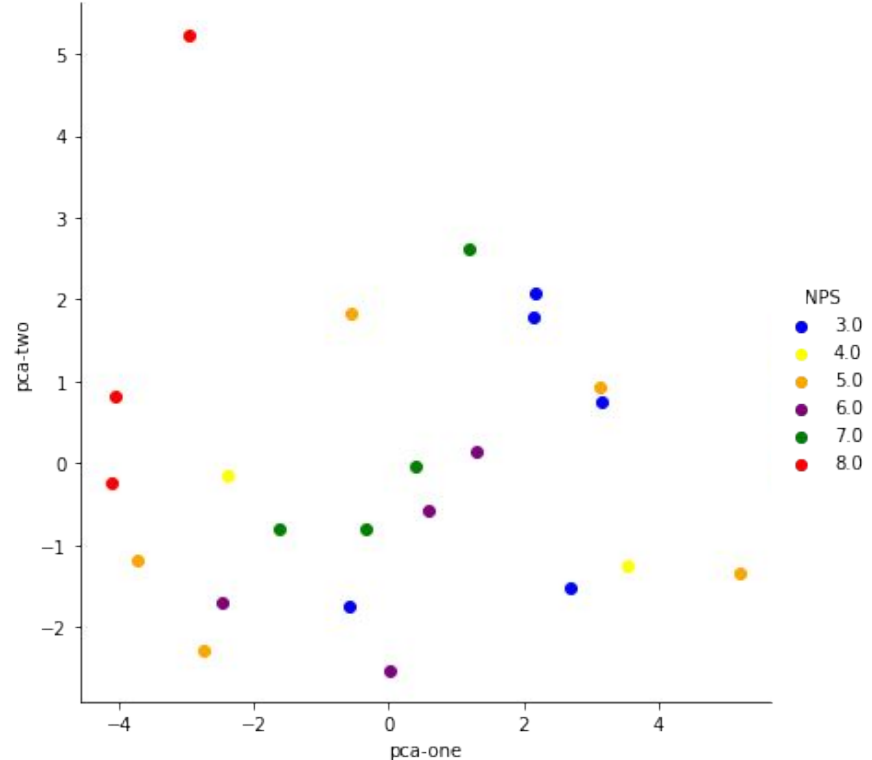
	Age	Département	Fréquence	Itinéraire	Appréciation- portail	Esthétisme	Itinéraire- satisfaction	Utilisation- fréquente	systeme- complexe	systeme- facile	...
0	-1.223073	-0.189382	1.256757	-0.308607	0.794101	-0.610847	0.556890	-0.578335	-1.100699	1.379461	...
1	-0.240954	-0.189382	0.145010	-0.308607	-2.015796	-2.015796	0.556890	-0.578335	-1.100699	-2.019925	...
2	-0.732014	-0.189382	1.256757	-0.308607	0.794101	0.794101	0.556890	0.371787	0.104828	0.246332	...
3	-1.152921	3.308441	1.256757	-0.308607	-0.610847	-0.610847	-1.272892	-0.578335	1.310356	-2.019925	...
4	-0.942467	-1.311325	-0.966736	-0.308607	0.794101	0.794101	-1.272892	-1.528456	1.310356	-0.886796	...
5	2.074038	-1.245328	-0.966736	-0.308607	-2.015796	0.794101	-1.272892	0.371787	0.104828	0.246332	...
6	2.074038	-0.189382	0.145010	-0.308607	0.794101	0.794101	-1.272892	0.371787	0.104828	-0.886796	...
7	-1.714132	-0.189382	-0.966736	-0.308607	-0.610847	0.794101	-2.187783	-1.528456	1.310356	0.246332	...
8	-0.030501	-0.189382	-0.966736	-0.308607	-0.610847	-2.015796	-0.358001	2.272029	1.310356	-0.886796	...
9	-0.872316	-0.189382	1.256757	-0.308607	0.794101	-0.610847	0.556890	-0.578335	-1.100699	-0.886796	...
10	-0.170803	-0.189382	1.256757	-0.308607	0.794101	0.794101	0.556890	1.321908	0.104828	0.246332	...

Dimensionality reduction (PCA)

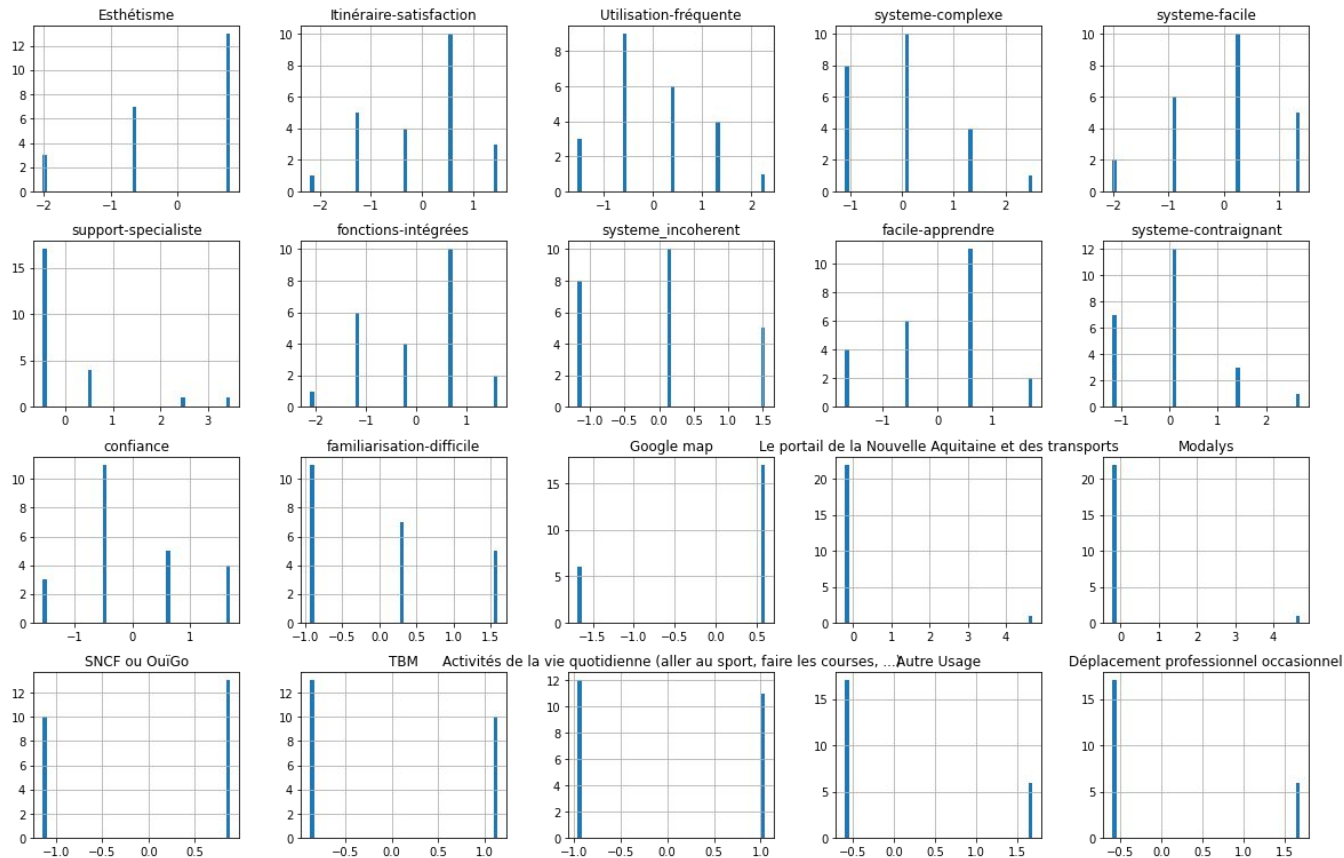
PCA = Principal Component Analysis

Why ?

- Reduce number of dimensions in data
- Find data patterns
- Visualize data



Data visualisation



Features
distribution

Machine learning models

- ❖ This is a supervised machine learning task. More specifically it's a classification problem.
- ❖ The machine learning models considered to train the dataset are:
 - Decision Tree
 - Multilayer Perceptrons
 - K-nearest neighbors
 - Gaussian naive bayes

Models Evaluation

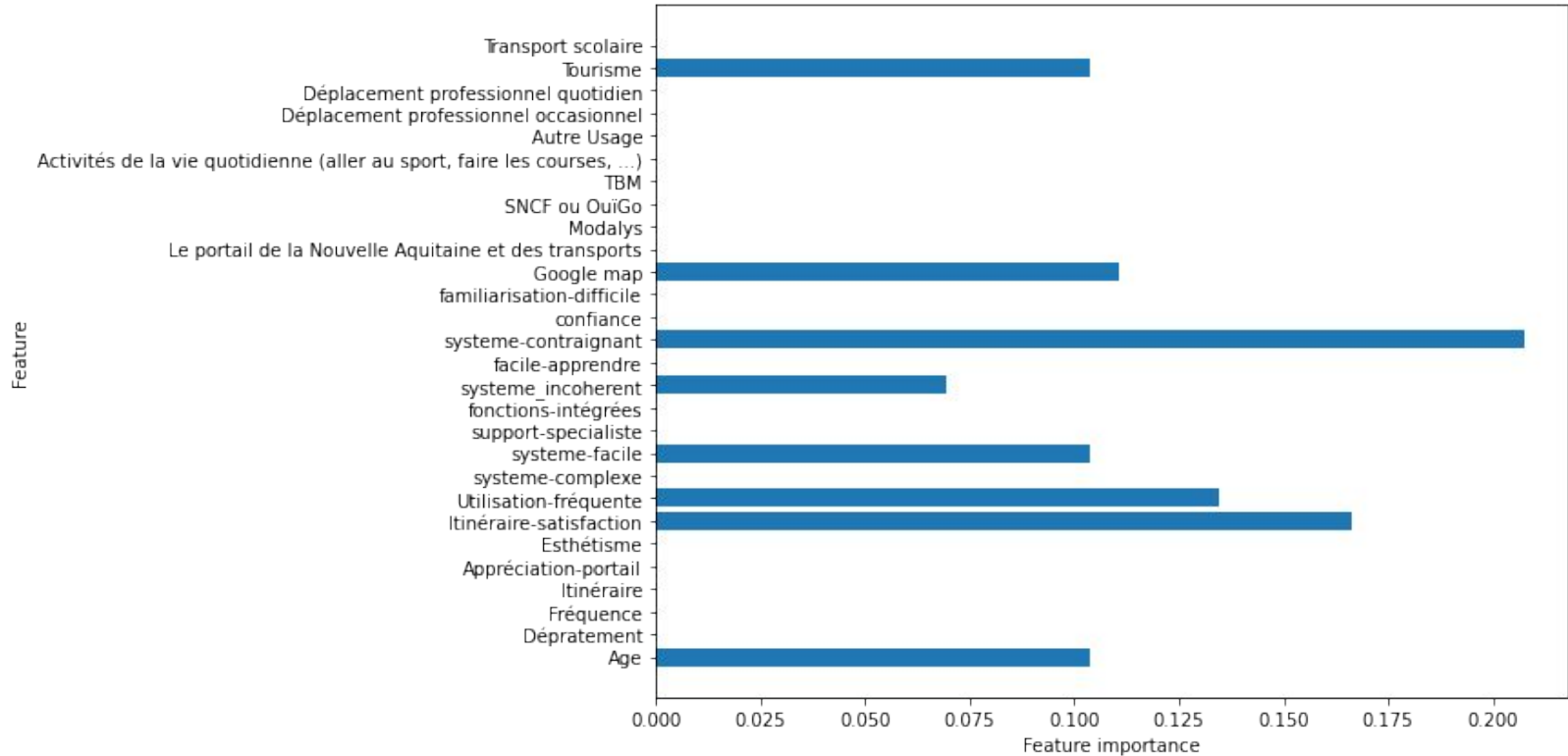
- The models are evaluated, and the considered metric is **accuracy**.

$$\text{Accuracy} = \frac{\text{Number of correct predictions}}{\text{Total number of predictions}}$$

- Below table shows the training and test dataset accuracy by the respective models:

ML Models	Train Accuracy	Test Accuracy
Decision Tree	0.83	0.26
Multilayer Perceptron	0.94	0.20
K-nearest neighbors	0.98	0.77
Gaussian Naive Bayes	0.72	0.22

Features Importance





03

Interface presentation

Interface features

- Descriptive statistics
 - Information on the uploaded data
- Cleaning and data analysis
 - Choices and how the data is cleaned
- Models analysis

NPS PREDICTION

Drag and Drop or Select Files

Data Overview

Processing

Model

NA-DATA.CSV

2021-11-05T21:41:31.585000

DATA INFORMATION

NUMBER OF INSTANCES

27

NUMBER OF FEATURES

29

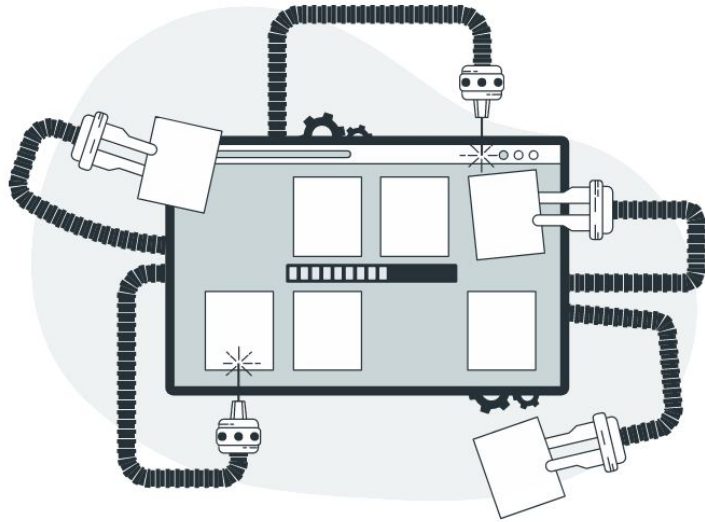
STATISTIC OF EACH FEATURE

Horodateur	
METRIC	VALUE
count	27
unique	27
freq	1

Number of NaN values in the dataset

NaN values

variable



04

Expected projection

Project deployment and enhancement

- Use time to forecast NPS.
- Deploy on Docker

Conclusion

- Cleaning processing a raw database
- Analysing and visualising data
- Creating an UX interface

Correlation between attributes

