

L'apprentissage automatique pour le traitement de données géospatiales

Forum TIC - 01/06/18

Raphaël Delhome - Oslandia

Présentation

- Spécialiste de l'Open Source, créée en 2009
- Systèmes d'Information Géographique, 3D, data
- 15 collaborateurs,
~1.3M€ de CA (2017)



Contenu de la présentation

Ce qu'il y aura

- Etudes de cas Oslandia
- Outils et méthodes

Contenu de la présentation

Ce qu'il y aura

- Etudes de cas Oslandia
- Outils et méthodes

Ce qu'il n'y aura pas

- Thématique biodiversité

Contenu de la présentation

Ce qu'il y aura

- Etudes de cas Oslandia
- Outils et méthodes

Ce qu'il n'y aura pas

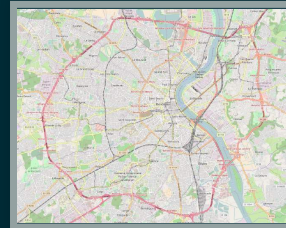
- Thématique biodiversité

Méthodes très génériques agnostiques au secteur d'activité

- Tant qu'il y a de la donnée géospatiale...

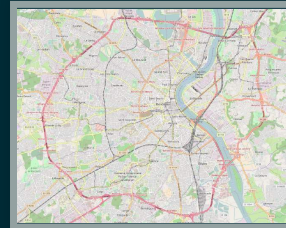
Exemples d'études Oslandia

- Evaluation de la qualité
des données
OpenStreetMap



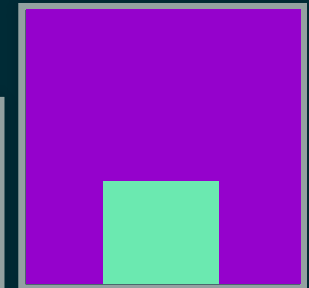
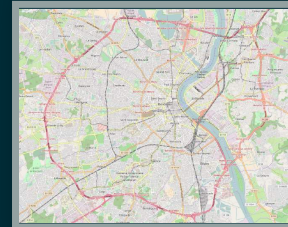
Exemples d'études Oslandia

- Evaluation de la qualité des données OpenStreetMap
- Traitement des données issues de système de vélos en libre-service



Exemples d'études Oslandia

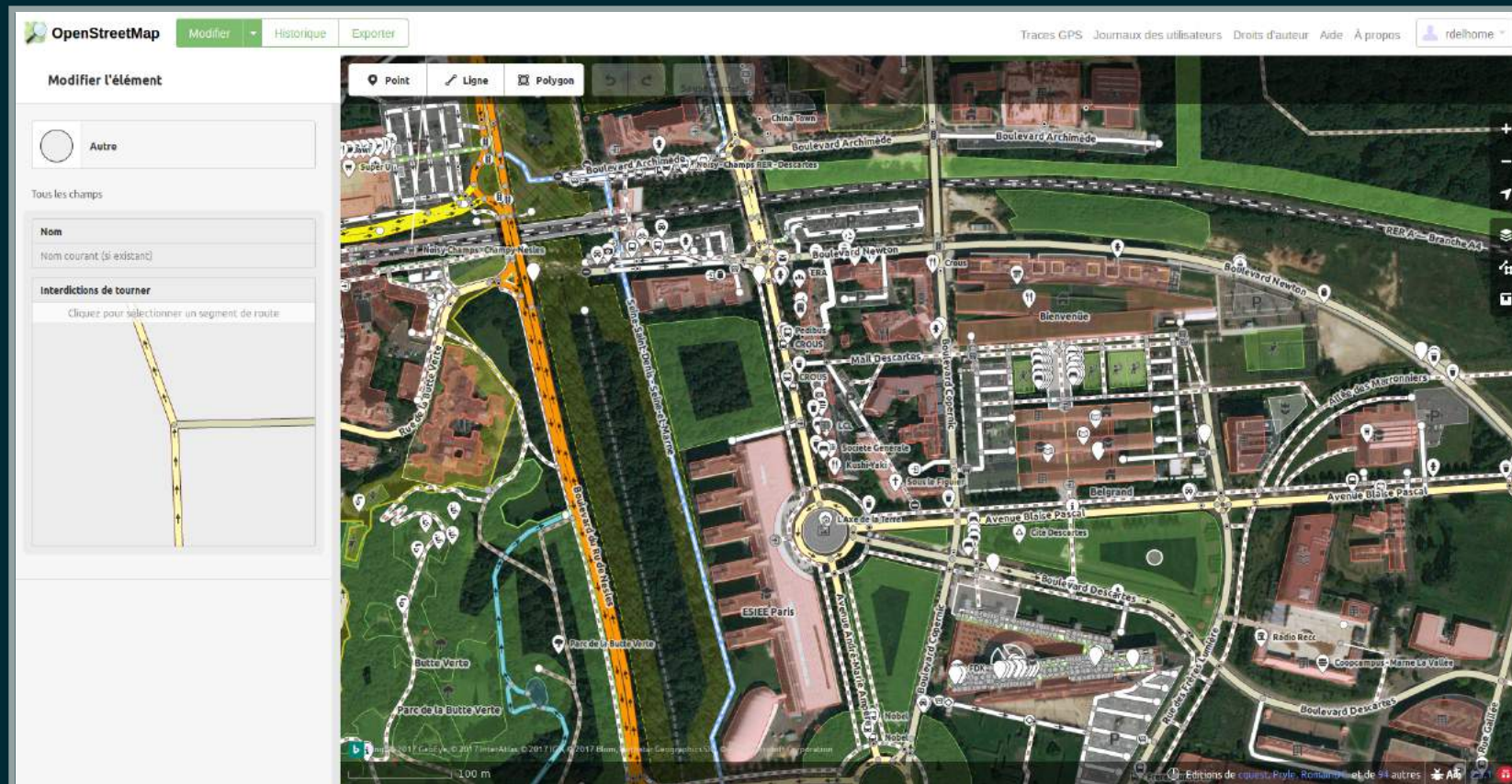
- Evaluation de la qualité des données OpenStreetMap
- Traitement des données issues de système de vélos en libre-service
- Analyse d'images par l'apprentissage profond



<http://data.oslandia.io>

OpenStreetMap (OSM)

OpenStreetMap (OSM)



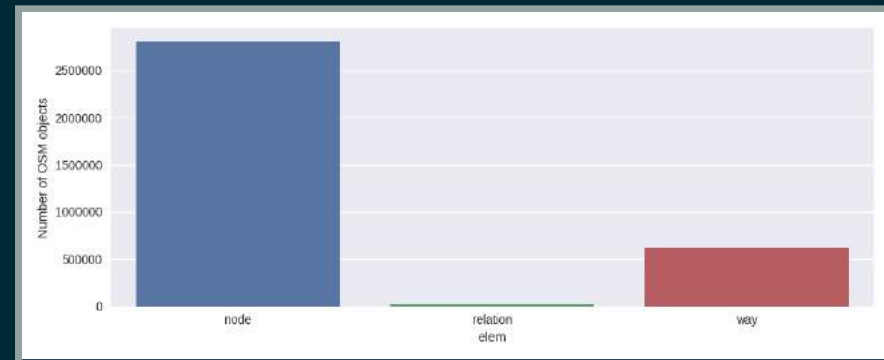
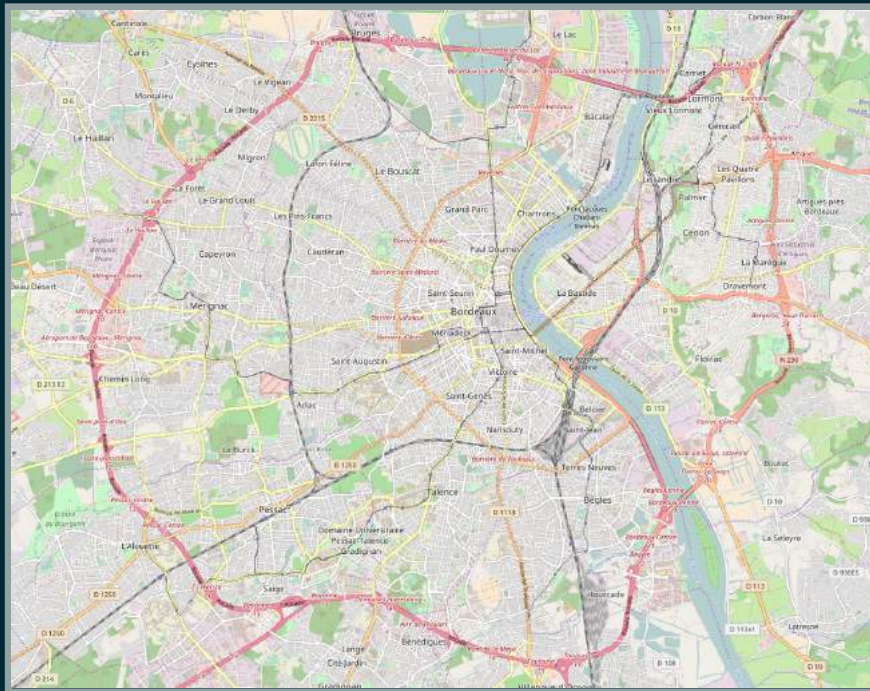
OSM: Cartographie collaborative

Qualité des données OSM

Question de recherche: *Peut-on évaluer la qualité des données?*

- Disponibilité d'une source de donnée alternative de référence
- Tout le monde *peut* contribuer, tout le monde *sait-il* contribuer ?
- Evaluer les objets OSM avec leurs métadonnées

Présentation de la donnée



```
<osm version="0.6" generator="CGImap 0.6.0 (6182 thorn-01.openstreetmap.org)"  
copyright="OpenStreetMap and contributors" attribution="http://www.openstreetmap.org/copyright/"  
license="http://opendatacommons.org/licenses/odbl/1-0/">  
  <node id="101439002" visible="true" version="6" changeset="25575532"  
    timestamp="2014-09-21T09:49:32Z" user="Geofreund1" uid="179581" lat="44.5153863" lon="0.8404732"/>  
</osm>
```

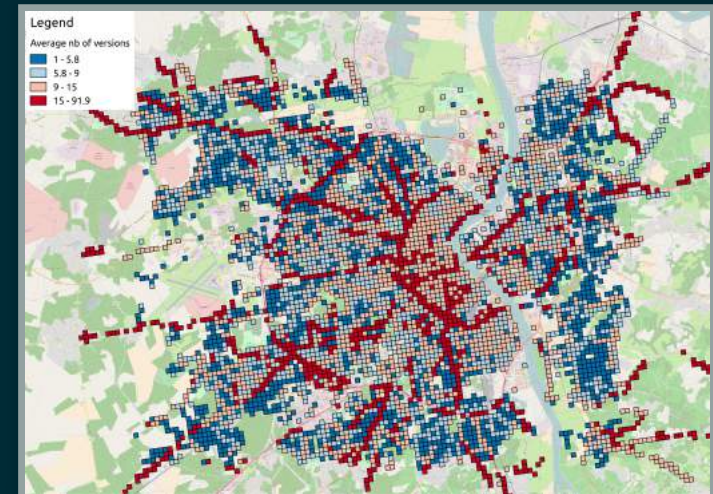
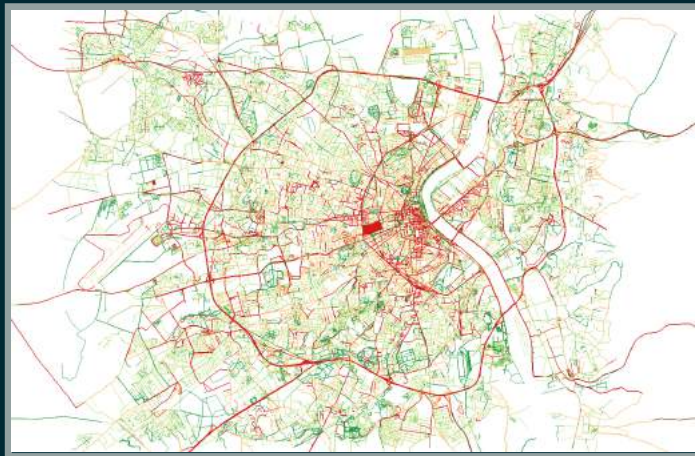
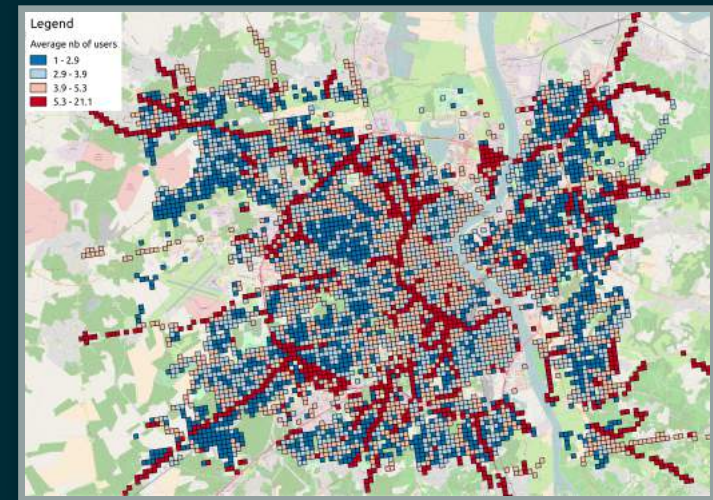
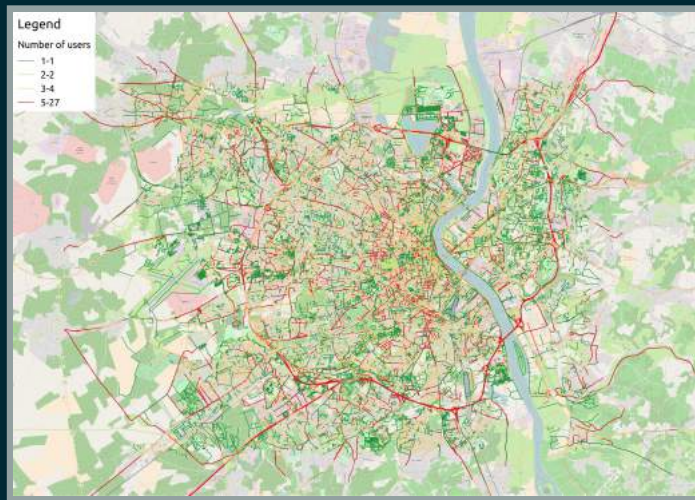


Extraction des métadonnées

En se focalisant sur les utilisateurs (*ex: user id*):

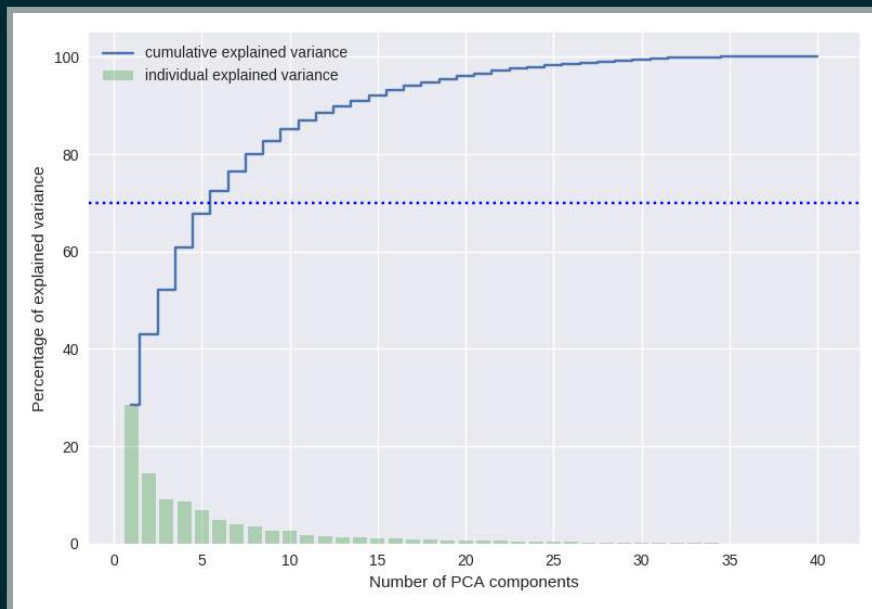
- Variables temporelles
 - sur OSM depuis 2136 jours, 1 jour d'activité
- Variables liées aux sets de modifications
 - 19 sets de modification, parmi lequel 1 autour Bordeaux
- Variables liées aux modifications
 - 4 modifications, (3 créations de noeuds, 1 amélioration de way), corrigé depuis
- Variables liées à l'éditeur utilisé
 - 14 modifs oavec Potlach, 3 avec iD, 2 avec JOSM

Cartographie à partir des métadonnées



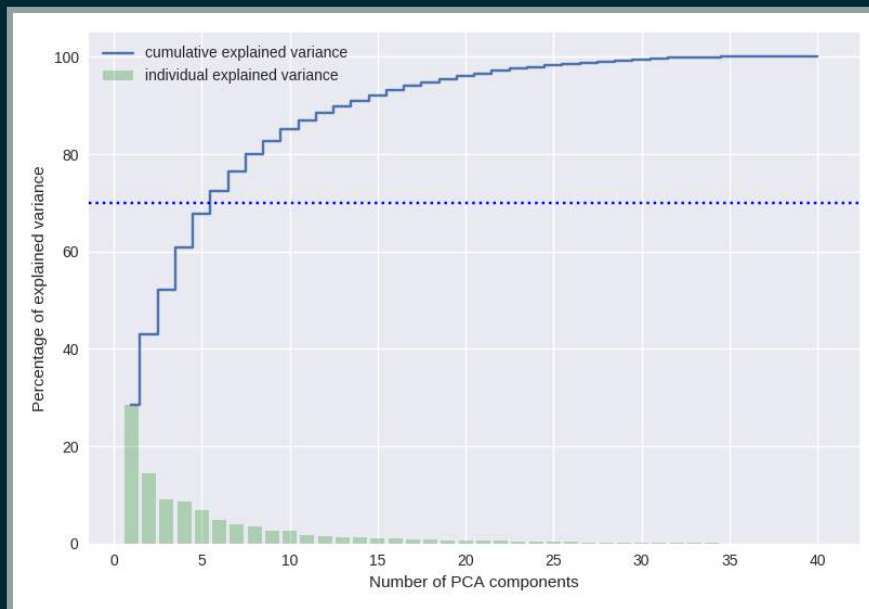
Réduction de dimension (Analyse en Composante Principale, ou ACP)

De 40 variables à juste une poignée : composantes les plus significatives



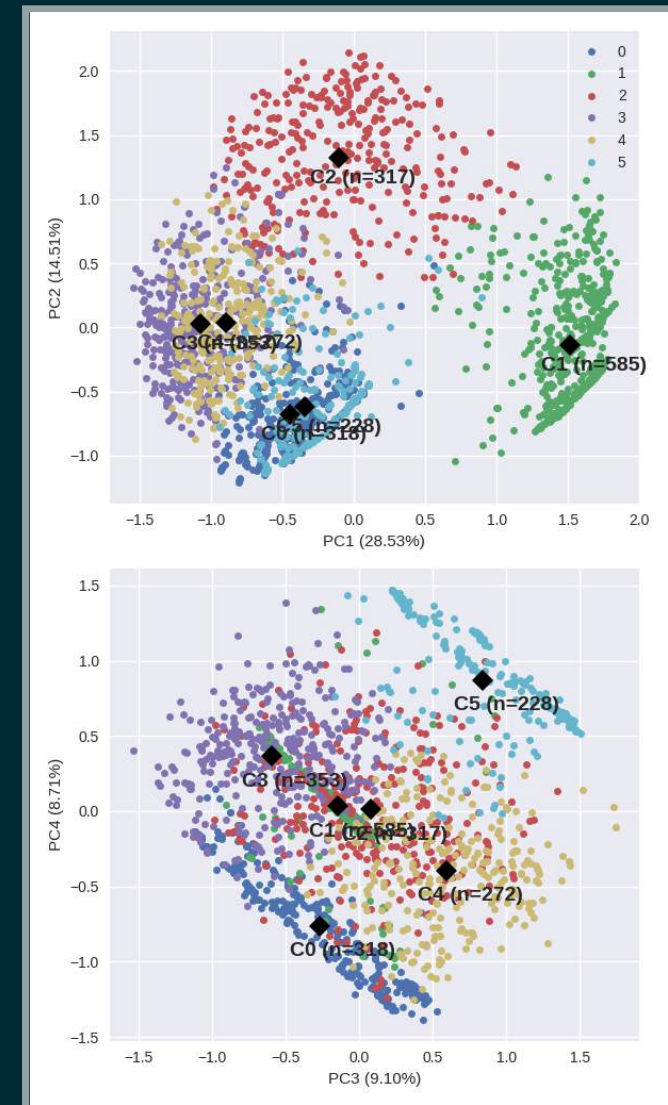
Réduction de dimension (Analyse en Composante Principale, ou ACP)

De 40 variables à juste une poignée : composantes les plus significatives



	PC1	PC2	PC3	PC4	PC5	PC6
lifespan	0.005	0.151	0.017	-0.014	-0.013	0.032
n_inscription_days	0.051	0.105	0.199	-0.169	-0.073	-0.015
n_activity_days	0.003	0.369	0.005	0.005	-0.081	0.111
u_chgset	-0.061	0.366	-0.068	0.050	-0.142	0.139
dmean_chgset	0.001	0.001	0.002	-0.001	-0.003	0.002
u_modif_byelem	0.089	0.161	-0.011	-0.011	-0.267	0.267
u_total_modif	-0.130	0.356	-0.031	-0.010	-0.099	-0.004
n_total_modif_node	-0.284	0.062	-0.194	-0.397	0.112	0.091
n_total_modif_way	-0.110	-0.032	0.303	0.352	0.037	-0.054
n_total_modif_relation	0.394	-0.030	-0.109	0.044	-0.149	-0.038
u_node_modif	-0.199	0.285	-0.085	-0.121	-0.013	-0.027
n_node_modif_cr	-0.251	0.052	-0.196	-0.181	-0.164	-0.492
n_node_modif_imp	-0.078	0.093	0.081	-0.182	0.233	0.590
n_node_modif_del	-0.027	0.051	-0.026	0.035	0.040	-0.002
n_node_modif_utd	-0.231	0.132	-0.370	0.077	0.235	-0.060
n_node_modif_cor	-0.107	0.037	0.226	-0.379	-0.091	0.071
n_node_modif_autocor	-0.021	0.031	0.002	-0.029	-0.041	0.096
u_way_modif	-0.174	0.293	0.074	0.147	-0.030	-0.127
n_way_modif_cr	-0.106	0.091	-0.101	-0.036	-0.058	-0.265
n_way_modif_imp	-0.203	0.141	0.391	0.388	0.042	0.060
n_way_modif_del	-0.014	0.024	-0.018	0.009	-0.000	0.001
n_way_modif_utd	-0.149	0.103	-0.145	0.309	0.133	-0.052
n_way_modif_cor	-0.142	0.101	0.398	0.045	-0.101	-0.180
n_way_modif_autocor	-0.040	0.065	0.023	0.009	-0.060	0.034
u_relation_modif	0.198	0.273	-0.078	0.043	-0.132	0.028
n_relation_modif_cr	-0.011	0.050	-0.020	0.010	-0.001	-0.026
n_relation_modif_imp	0.383	0.241	-0.123	0.072	-0.174	-0.019
n_relation_modif_del	-0.001	0.006	-0.004	0.004	0.005	-0.003
n_relation_modif_utd	-0.011	0.060	-0.057	0.049	0.009	0.002
n_relation_modif_cor	0.310	0.122	-0.078	0.040	-0.051	-0.153
n_relation_modif_autocor	0.073	0.115	-0.012	-0.003	-0.128	0.104
u_total_chgset	0.159	0.205	0.030	-0.025	0.250	-0.050
p_local_chgset	-0.230	-0.127	-0.054	0.078	-0.360	0.106
n_total_chgset_id	-0.147	-0.126	-0.362	0.348	-0.148	0.224
n_total_chgset_josm	0.178	0.198	0.109	-0.060	0.491	-0.187
n_total_chgset_maps.me_android	-0.008	-0.020	-0.009	-0.012	0.012	0.010
n_total_chgset_maps.me_ios	-0.006	-0.010	-0.009	-0.002	0.012	0.005
n_total_chgset_other	-0.002	0.002	0.014	-0.030	0.020	0.021
n_total_chgset_potlatch	-0.019	-0.044	0.210	-0.185	-0.377	-0.109
n_total_chgset_unknown	0.004	-0.001	0.048	-0.058	-0.010	0.036

Classification des utilisateurs



Classification des utilisateurs

Apprentissage non-supervisé: K-means

G0 (n=318) Contributeurs peu expérimentés (noeuds)

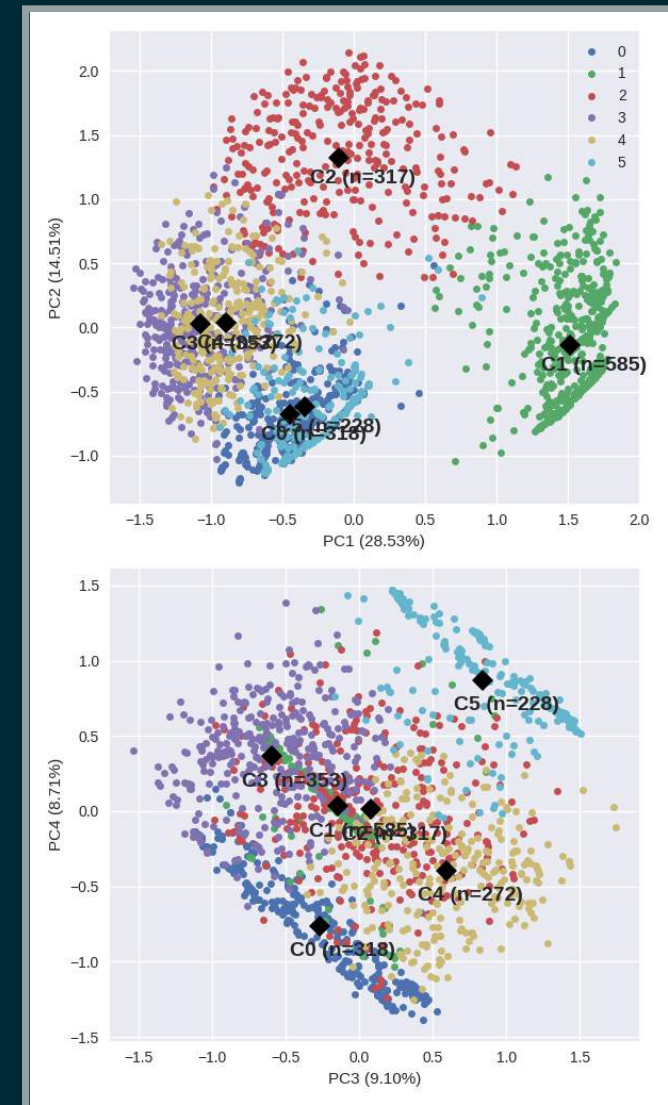
G1 (n=585) Utilisateurs expérimentés (relations), peu de contributions locales

G2 (n=317) Experts OSM polyvalents

G3 (n=353) Nouveaux contributeurs

G4 (n=272) Anciens utilisateurs (une seule contribution)

G5 (n=228) Spécialistes des ways, mais contributions rares

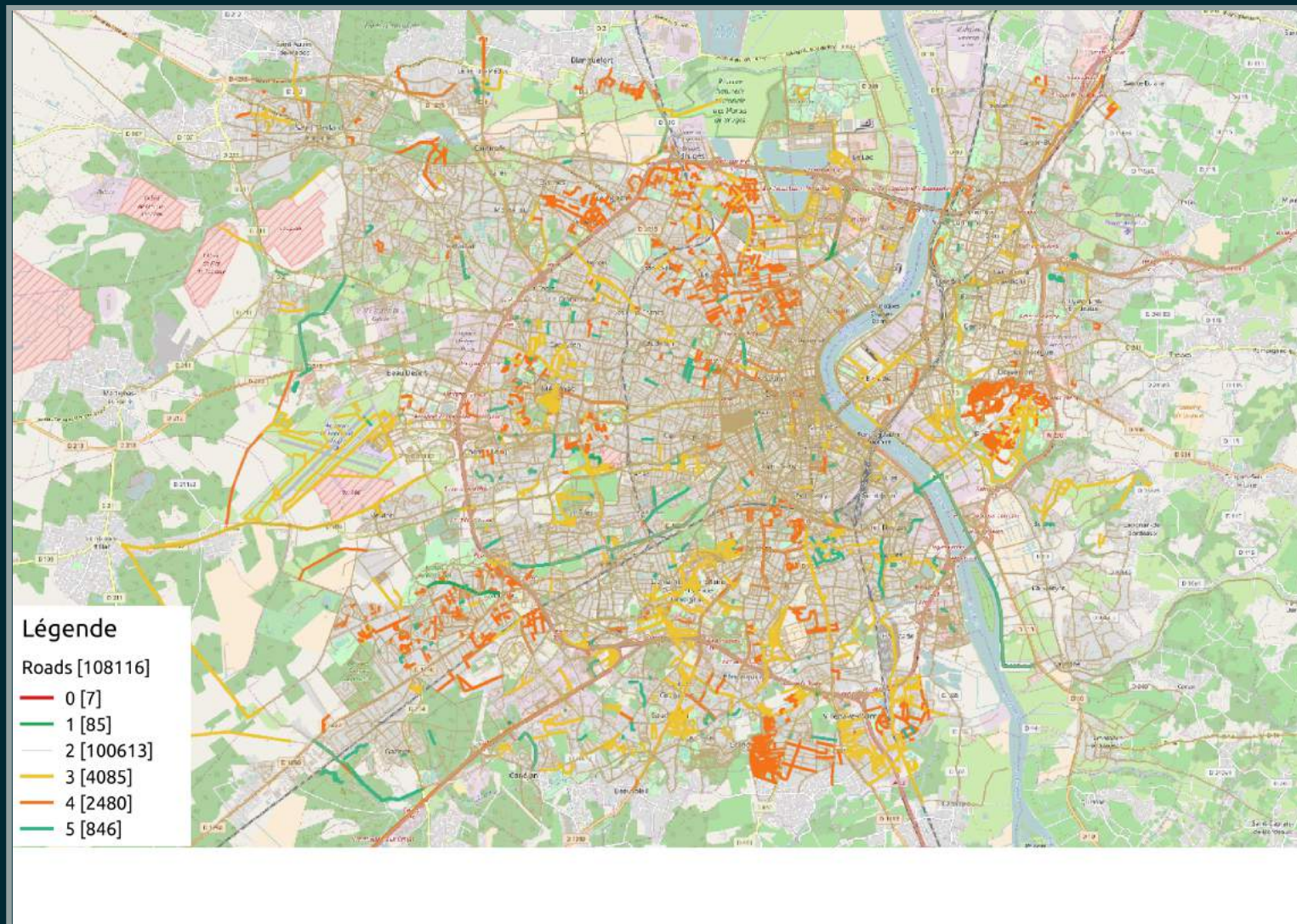


Utilisation de la classification

Exemple: Etiquetage des routes -> contributeurs expérimentés = objets OSM de bonne qualité

Groupes	Quantité	Objets
G0	318 (15.3%)	6690 (0.2%)
G1	585 (28.2%)	164 (0.0%)
G2	317 (15.3%)	2670391 (96.7%)
G3	353 (17.0%)	23287 (0.8%)
G4	272 (13.1%)	9993 (0.4%)
G5	228 (11.0%)	50474 (1.8%)

Cartographie





Conclusion

Conclusion

- Caractériser les métadonnées OSM avec de l'apprentissage non-supervisé
- Evaluation de la qualité des données OSM: l'exemple des routes à Bordeaux
- Comment aller plus loin?
 - complétude des métadonnées: tous les utilisateurs, toutes les contributions (pas seulement le local)
 - comparaison avec une vérité terrain: données IGN, orthophotos Bing...?

Etude des systèmes de vélos en libre-service

Services de vélos partagés



- Vélos en libre-service
- Location de courte durée
- Stations et disponibilités

Velo'v (Lyon)

Données géospatiales ouvertes



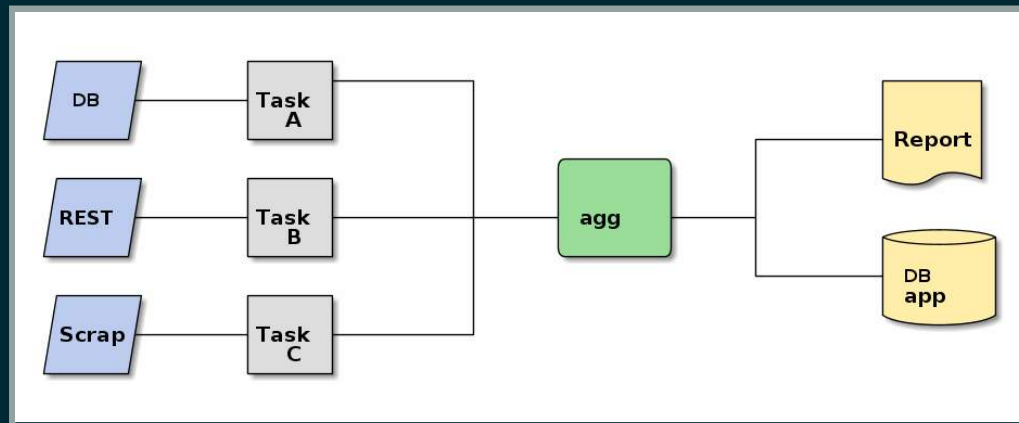
Data GrandLyon



Opendata Bordeaux

Gérer la donnée

- Construire un pipeline de données avec Luigi
- Obtenir, transformer et stocker la donnée
 - récolte toute les 10 minutes (j son, xml, shp)
 - stockage en base (postgresql, postgis)
 - Ingénierie de la donnée et machine learning



Classification des stations (K-means)

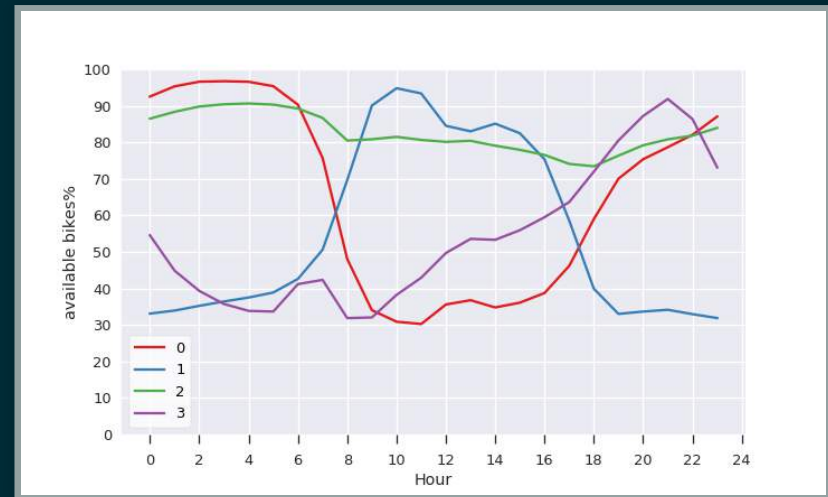
Classer les stations à partir de leur utilisation par les clients

- Inspiré d'un travail similaire à Dublin
- 1 profil = 1 individu
- Grouper les individus proches
- Dédire des profils "types"

Classification des stations (K-means)

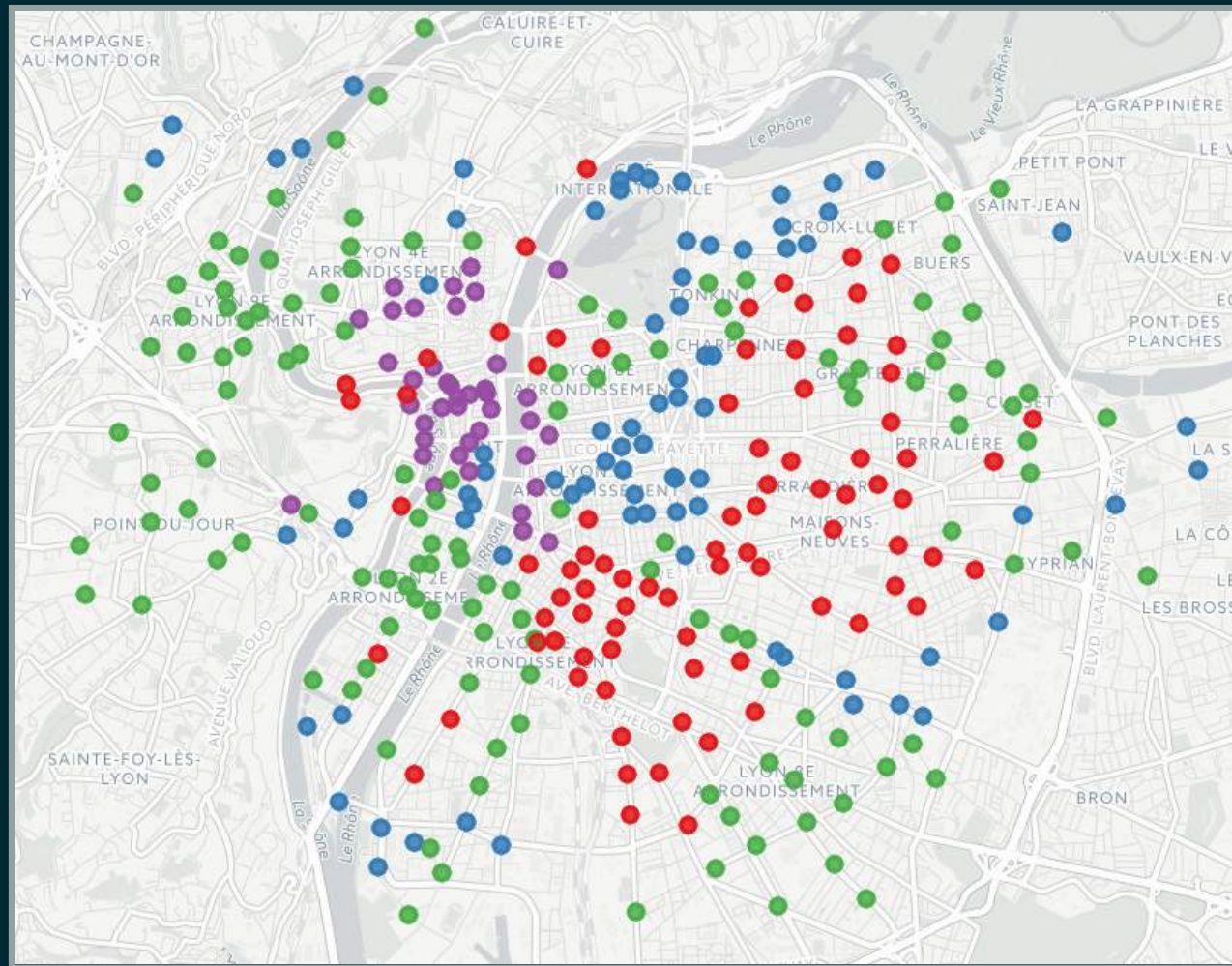
Classer les stations à partir de leur utilisation par les clients

- Inspiré d'un travail similaire à Dublin
- 1 profil = 1 individu
- Grouper les individus proches
- Dédire des profils "types"



Exemple avec 4 clusters

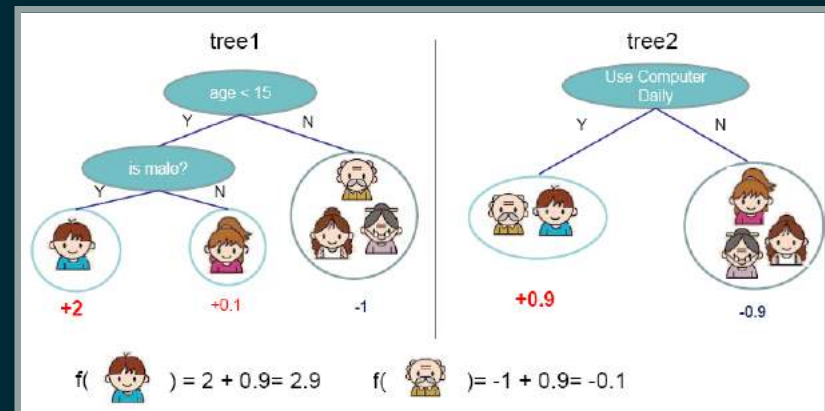
Cartographie des stations



Prédiction de la disponibilité des vélos

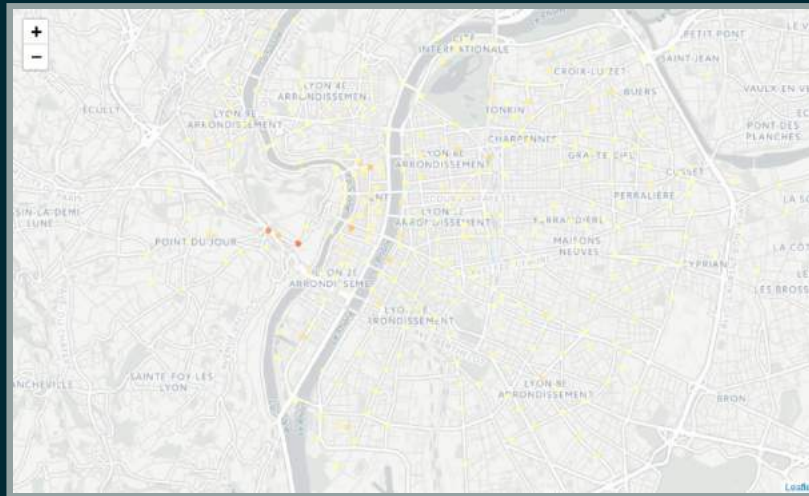
Apprentissage supervisé
(méthode XGBoost) :

- prédire Y (proba. de disponibilité à $H+1$)
- à partir de X (heure, jour, vélos disponibles à H , ...)

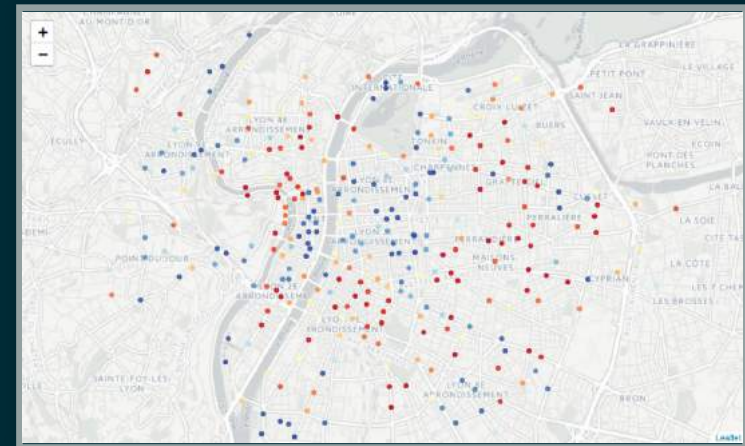
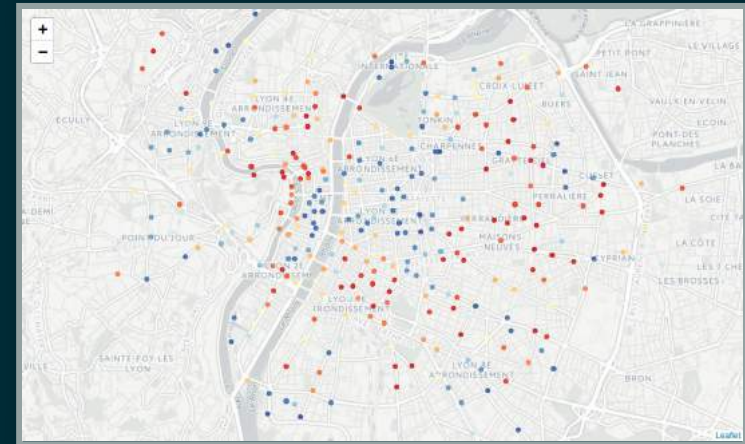


Résultats

Sans ingénierie particulière, RMSE = 0.095



- gauche: erreurs
- haut: prédictions
- bas: réalité





Conclusion

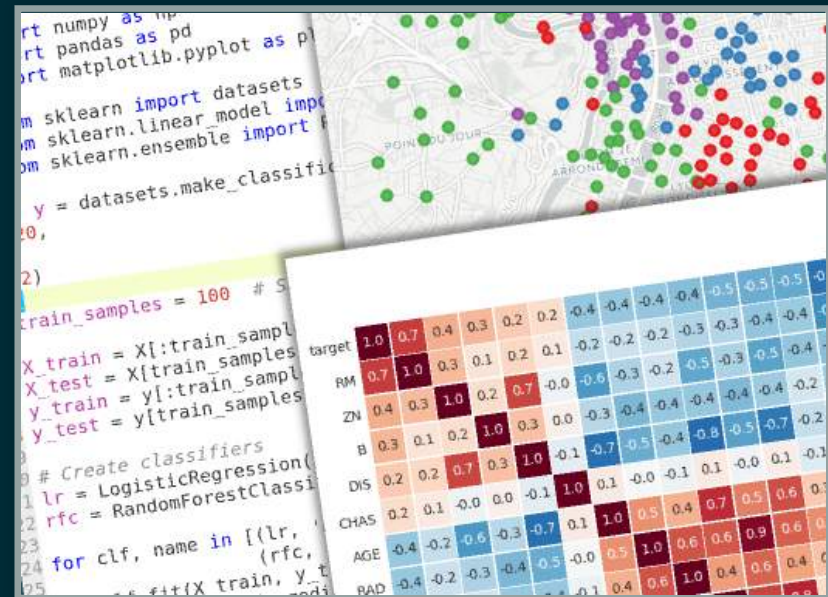
Conclusion

- Traiter des questions de recherche simples avec un jeu de données géospatiales ouvert
- Construire un pipeline de données
- Production d'une **application web** pour visualiser la donnée
- Ouverture : "online learning", ou continuer à agréger de la donnée et modifier les prédictions en direct

Analyse d'image par l'apprentissage profond

Contexte du projet

- Développement des activités data
- Démocratisation de l'apprentissage profond
- Maturité technique des paquets Python



Objectifs



- Analyse automatique des données images
- Montée en compétence interne

Types d'analyse

- *Classification*: déterminer le type d'objet représenté sur une image

Types d'analyse

- *Classification*: déterminer le type d'objet représenté sur une image
- *Détection d'objets*: vérifier si des objets sont présents sur une image

Types d'analyse

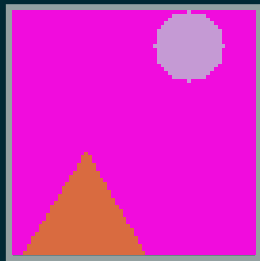
- *Classification*: déterminer le type d'objet représenté sur une image
- *Détection d'objets*: vérifier si des objets sont présents sur une image
- *Localisation d'objets*: localiser chaque objet sur une image

Types d'analyse

- *Classification*: déterminer le type d'objet représenté sur une image
- *Détection d'objets*: vérifier si des objets sont présents sur une image
- *Localisation d'objets*: localiser chaque objet sur une image
- *Segmentation sémantique*: étiqueter chaque pixel par l'objet auquel il appartient

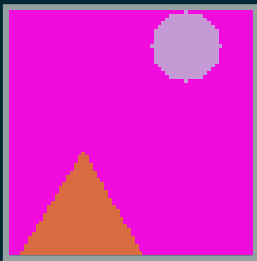
Jeux de données

- *Shapes*: images générées aléatoirement avec un carré, un cercle et/ou un triangle

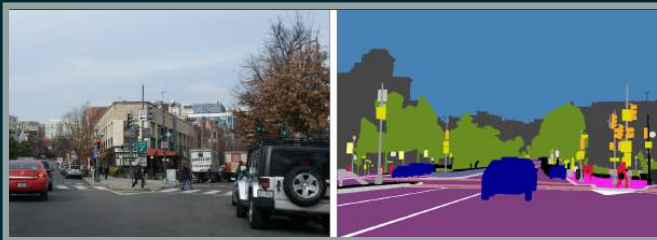


Jeux de données

- *Shapes*: images générées aléatoirement avec un carré, un cercle et/ou un triangle



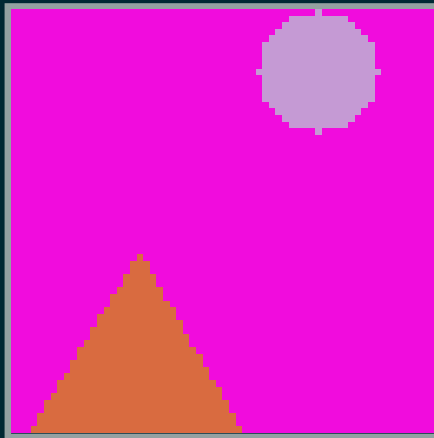
- *Mapillary*: images de scènes urbaines, avec 66 types d'objet



Détection d'objets

Inputs Lots d'images [*dimension = (lot, pixel, pixel, canaux)*]

Outputs Lot de vecteurs d'étiquettes [*dimension = (lot, étiquettes)*]



- L'image contient: **cercle - triangle**
- L'image ne contient pas: **carré**
- Résultat: [0, 1, 1]

Algorithmes

- Réseau de neurone convolutionnel "naïf":
 - quelques blocs de convolutions au début
 - quelques couches denses à la fin
- **VGG16**: un réseau *plus gros* avec plus de couches
- **Inception**: un réseau *plus intelligent* avec des blocs innovants

Segmentation sémantique

Inputs Lots d'images [*dimension* = (lot, pixels, pixels, canaux)]

Outputs Lots de vecteurs d'étiquettes [*dimension* = (lot, pixels, pixels, étiquettes)]



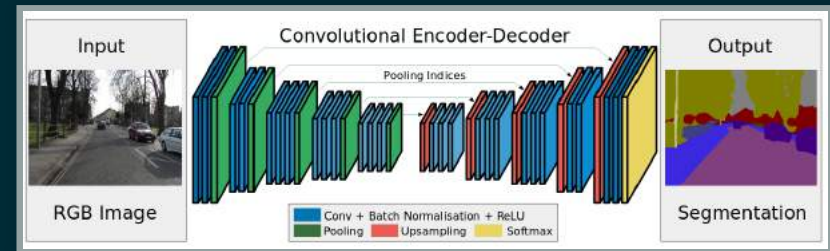
- Résultat: 1 si le pixel (p, p) de l'image i appartient à un objet de la classe c
- Construit une version "étiquetée" de chaque image

Algorithmes

- Fully Convolutional Network: ne pas utiliser de couche dense (tout convolutionnel)

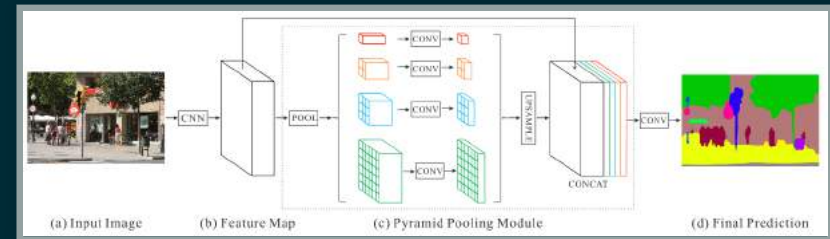
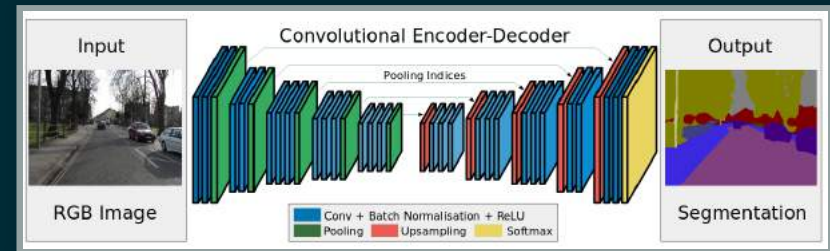
Algorithmes

- Fully Convolutional Network: ne pas utiliser de couche dense (tout convolutionnel)
- Encoder-decoder architecture



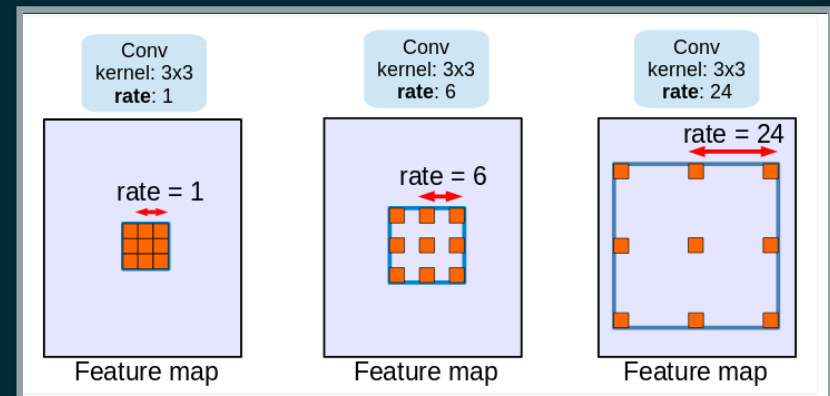
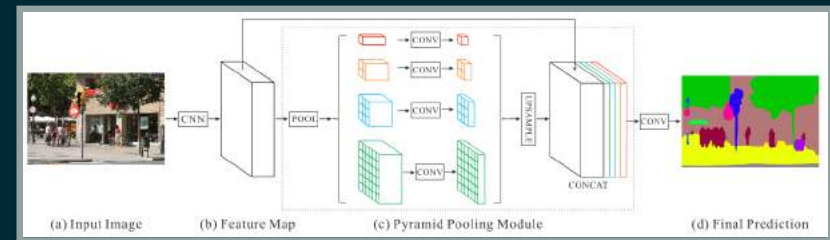
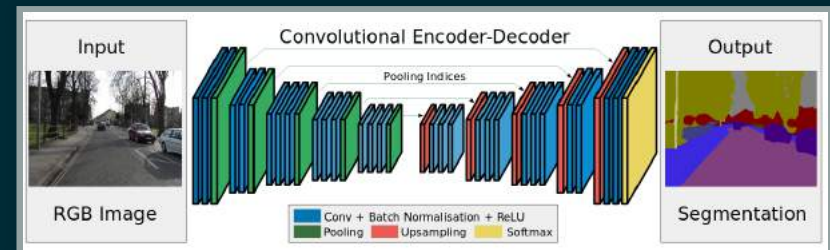
Algorithmes

- **Fully Convolutional Network:** ne pas utiliser de couche dense (tout convolutionnel)
- **Encoder-decoder architecture**
- **PSPNet:** structure pyramidale



Algorithmes

- Fully Convolutional Network: ne pas utiliser de couche dense (tout convolutionnel)
- Encoder-decoder architecture
- PSPNet: structure pyramidale
- Dilated/Atrous convolutions



Résultats

Parcourir...

Aucun fichier sélectionné.

Upload



- animal-bird: 0%
- construction-barrier-curb: 84%
- construction-flat-bike-lane: 100%
- construction-structure-bridge: 100%
- human-person: 8%
- marking-crosswalk-zebra: 100%
- nature-mountain: 100%
- object-banner: 100%
- object-support-traffic-sign-frame: 100%
- object-vehicle-bicycle: 100%
- void-car-mount: 100%

Predict

- Production d'une petite application web pour la détection d'objets
- Segmentation sémantique bientôt disponible...



Conclusion

Conclusion

- L'apprentissage profond chez Oslandia : sujet récent
- Etiquetage d'images, agnosticisme vis-à-vis du secteur d'application
- Perspective: lier segmentation sémantique et 3D



Merci de votre attention!

Questions?

raphael.delhome@oslandia.com

Pour aller plus loin :

- <http://oslandia.com/en/blog/>
- <https://github.com/Oslandia>
- <http://data.oslandia.io>