***République Algérienne Démocratique et Populaire***

***Ministère de l’Enseignement Supérieur et de la Recherche Scientifique***

***Université des Sciences et de la Technologie Houari Boumediene***

***Faculté d’Informatique***

****

***Segmentation géospatiale selon le type d’infrastructures et Bâtiments (Cas : Ville d’Alger)***

***Réalisé par :***

TAOUINT ABDERRAHMANE

# 

Table des matières

[Introduction Générale 3](#_Toc198574576)

[Chapitre 1 : État de l’art / Étude de l’existant 5](#_Toc198574577)

[1.1 Introduction et problématique 5](#_Toc198574578)

[1.2 A propos du fournisseur téléphonique DJEZZY 5](#_Toc198574579)

[1.3 Organisation actuelle du suivi opérationnel 6](#_Toc198574580)

[1.4 Autres usages internes pour Djezzy 6](#_Toc198574581)

[1.5 Problèmes rencontrés dans le suivi technique 7](#_Toc198574582)

[1.5.1 Conséquences concrètes pour l’entreprise 7](#_Toc198574583)

[1.5.2 Difficultés pour les utilisateurs finaux 7](#_Toc198574584)

[1.6 Méthodes existantes d’analyse de données et de cartographie 7](#_Toc198574585)

[1.6.1 Supervision classique 7](#_Toc198574586)

[1.6.2 Systèmes d’Information Géographique (SIG) 7](#_Toc198574587)

[1.7 Solution envisagée 7](#_Toc198574588)

[1.7.1 Objectifs de la solution 8](#_Toc198574589)

[1.7.2 Description générale du système 8](#_Toc198574590)

[Conclusion du chapitre 8](#_Toc198574591)

[2. Chapitre 2 : Analyse et conception de la solution 9](#_Toc198574592)

[2.1 Traitement des données raster 9](#_Toc198574593)

[2.1.1 Importance cruciale de la qualité des données raster pour le traitement spatial 9](#_Toc198574594)

[2.1.2 Sources et nature des données 10](#_Toc198574595)

[2.2 Pipeline d'analyse et de segmentation 10](#_Toc198574596)

[2.2.1 Filtrage spatial avec HDBSCAN 10](#_Toc198574597)

[2.2.2 Regroupement par clustering KMeans (MiniBatch) 11](#_Toc198574598)

[2.2.3 Attribution des labels et reconstruction géographique 11](#_Toc198574599)

[2.3 Visualisation cartographique 12](#_Toc198574600)

[2.3.1 Choix de l'outil : Folium 12](#_Toc198574601)

[2.3.2 Structure de la carte finale 12](#_Toc198574602)

[2.3.3 Export et intégration 12](#_Toc198574603)

[2.4 Architecture globale du pipeline 12](#_Toc198574604)

[13](#_Toc198574605)

[2.5 Contraintes techniques rencontrées 13](#_Toc198574606)

[Conclusion du chapitre 13](#_Toc198574607)

# Introduction Générale

Nous vivons une ère de transformation numérique profonde, dans laquelle les technologies de l’information et de la communication jouent un rôle clé dans l’évolution des entreprises. Ces technologies permettent d’optimiser les processus, d’automatiser les tâches répétitives et de prendre des décisions plus rapidement grâce à l’exploitation intelligente des données. Parmi les secteurs les plus concernés par ces avancées, celui des télécommunications occupe une place stratégique.

En Algérie, Djezzy fait partie des principaux opérateurs mobiles. Face à l’augmentation constante du nombre d’utilisateurs, de la demande en données, et à la complexité croissante des réseaux (3G, 4G, voire 5G), Djezzy est confronté à plusieurs défis majeurs : assurer une couverture optimale sur l’ensemble du territoire, identifier rapidement les zones mal desservies, anticiper les saturations, et optimiser le déploiement de ses infrastructures.

Traditionnellement, ces opérations nécessitent une intervention humaine, des relevés sur terrain, et une analyse manuelle des données réseau. Ce processus est long, coûteux et peu évolutif. Par ailleurs, les outils actuellement utilisés offrent une visibilité limitée sur les zones critiques du réseau, en particulier dans les régions à forte densité de population ou à topographie complexe.

C’est dans ce contexte que s’inscrit notre projet de fin de cycle, réalisé en collaboration avec Djezzy. Il vise à proposer une solution intelligente et automatisée d’analyse spatiale du réseau mobile, en exploitant des données géographiques fournies sous forme de couches raster (fichiers .tif). À partir de ces données, notre application permet d’identifier les zones stratégiques (fortement peuplées, saturées ou sensibles), de les regrouper selon leur profil, et de les afficher de façon claire et interactive sur une carte dynamique.

Cette solution repose sur un enchaînement méthodologique rigoureux : traitement et nettoyage des données raster, détection automatique des zones significatives, regroupement en clusters homogènes, puis visualisation géographique à l’aide de technologies web. L’approche retenue se veut à la fois automatisée, interactive et évolutive. Elle s’adresse principalement aux ingénieurs réseau et aux décideurs de Djezzy qui souhaitent disposer d’un outil d’aide à la décision, capable de rendre visible, en quelques secondes, des informations précieuses sur le terrain.

L’originalité de notre projet réside dans l’intégration de l’analyse spatiale automatisée à une interface cartographique moderne. Il permet non seulement de localiser les zones d’intérêt, mais aussi de les comparer, de suivre leur évolution et d’appuyer la planification stratégique des déploiements ou des interventions techniques.

Le rapport est structuré de la manière suivante :

**Introduction** – Présentation du contexte du projet, des problématiques liées à l’analyse de la couverture réseau mobile et des objectifs visés à travers le développement d’un outil d’analyse spatiale intelligent et interactif.

**Chapitre 1 : État de l’art et étude de l’existant** – Ce chapitre présente l’état actuel de la gestion du réseau mobile chez Djezzy, les outils utilisés, les difficultés rencontrées, ainsi qu’une revue des approches existantes en matière d’analyse spatiale et de visualisation cartographique.

**Chapitre 2 : Présentation de la solution proposée** – Ce chapitre décrit la solution imaginée pour répondre aux besoins identifiés. Il expose ses fonctionnalités principales, les choix technologiques retenus et les apports attendus en matière d’automatisation, de visibilité réseau et d’aide à la décision.

**Chapitre 3 : Réalisation pratique et analyse des résultats** – Ce chapitre présente l’environnement de développement utilisé, les différentes étapes de conception et d’implémentation de l’outil, ainsi qu’une première analyse des résultats obtenus à partir des données fournies.

**Conclusion et perspectives** – Cette dernière partie dresse un bilan général du projet, met en évidence ses apports concrets et ses limites, et propose des pistes d’amélioration ou d’évolution pour un déploiement réel à grande échelle.

# Chapitre 1 : État de l’art / Étude de l’existant

## Introduction et problématique

Avant de proposer une solution technique, il est important de comprendre le contexte actuel dans lequel opère Djezzy, ainsi que les contraintes spécifiques liées à l’analyse des zones d’activité mobile sur le territoire national. Face à l’augmentation constante du nombre d’abonnés, à la diversité des zones géographiques couvertes (urbaines, rurales, enclavées) et à l’évolution rapide des technologies (passage vers la 4G et 5G), le besoin de disposer d’outils intelligents d’aide à l’évaluation des performances territoriales devient un enjeu stratégique.

Malgré l’utilisation d’outils classiques de supervision, l’identification des zones mal desservies ou fortement sollicitées reste difficile, souvent lente, et nécessite une forte intervention humaine. Le manque de visualisation claire, combiné à une analyse partiellement manuelle des données techniques, limite la réactivité des équipes terrain et la prise de décision stratégique.

En résumé, la problématique à laquelle notre projet tente d’apporter une réponse est la suivante :

**« Comment identifier, regrouper et visualiser automatiquement les zones critiques liées à l’activité mobile, afin de permettre une meilleure prise de décision et une optimisation continue de la qualité de service**

## A propos du fournisseur téléphonique DJEZZY

L'opérateur de télécommunications Algérien **Djezzy**, fondé en **juillet 2001**, est leader dans le secteur de la téléphonie mobile en Algérie. Avec une base de plus de **15 millions d'abonnés en septembre 2023**, Djezzy offre une gamme complète de services, allant du **prépayé au post-payé**, en passant par la **Data**, les **services à valeur ajoutée** et le **SUT**. Djezzy s'est toujours distinguée par son engagement envers l'innovation et la qualité de service, visant aujourd’hui à se positionner comme un acteur technologique de premier plan dans le pays.

Depuis **juillet 2022**, un tournant majeur s'est opéré avec le **transfert de la quasi-totalité des actions de VEON au profit du FNI-Fonds National d'Investissement**. En devenant une **Entreprise Publique Economique**, Djezzy est désormais détenue à hauteur de **96,57% par le FNI** qui joue un rôle significatif dans le développement socio-économique du pays. L'entreprise s'est fixée pour objectif d'étendre son réseau commercial sur l'ensemble du territoire national, notamment dans le Grand Sud, en ouvrant des boutiques dans dix nouvelles wilayas.

Djezzy a déjà accompli des réalisations significatives en matière de connectivité, couvrant **95 % de la population** à travers l'Algérie et déployant avec succès des services **3G** et **4G** dans les **58 wilayas**. Ces jalons clés dans son histoire, tels que l'octroi des licences 2G, 3G et 4G, illustrent son engagement constant envers l'expansion de la technologie.

L'ambition de Djezzy ne s'arrête pas là. En tant qu'entreprise publique économique, elle vise à devenir un acteur technologique majeur en Algérie. Son objectif est de favoriser le développement de la connectivité et de contribuer activement à la digitalisation du pays. Forte de son expérience et de son leadership, Djezzy s'efforce de façonner l'avenir technologique de l'Algérie en proposant des solutions innovantes et en garantissant un accès de qualité à la communication et à la connectivité à travers tout le pays.

## 1.3 Organisation actuelle du suivi opérationnel

Le suivi des activités techniques chez Djezzy repose actuellement sur une combinaison d’outils logiciels internes, de tableaux de bord analytiques, et d’interventions humaines. Les données collectées proviennent de capteurs installés sur les antennes, de logs système et des remontées clients.

Cependant, l’analyse spatiale de ces données est souvent fragmentée. Les équipes techniques utilisent des cartes statiques ou des outils complexes pour localiser les zones à faible signal, les surcharges ou les incidents. Le traitement est en partie automatisé, mais nécessite encore une expertise forte pour interpréter les résultats et décider des actions à entreprendre.

## 1.4 Autres usages internes pour Djezzy

Au-delà de l’analyse de couverture, ces mêmes mécanismes peuvent être appliqués à :

* L’optimisation du déploiement d’infrastructures
* La détection des zones de congestion
* Le ciblage géographique marketing
* La planification d’interventions techniques
* Le suivi des plaintes clients
* L’analyse prédictive

## 1.5 Problèmes rencontrés dans le suivi technique

Malgré les efforts déployés, plusieurs problèmes subsistent :

* **Absence de visualisation intuitive** : les outils actuels sont souvent techniques et non cartographiques.
* **Manque d’automatisation dans l’analyse** : la détection des zones critiques repose sur des seuils ou sur des rapports manuels.
* **Temps de réaction élevé** : en cas de problème, la localisation précise prend du temps.
* **Pas de regroupement des zones similaires** : ce qui complique la planification et la priorisation des interventions.

### 1.5.1 Conséquences concrètes pour l’entreprise

* **Perte d'efficacité opérationnelle** : difficulté à agir rapidement sur les zones sensibles.
* **Charge supplémentaire pour les équipes techniques** : analyse manuelle et surveillance permanente.
* **Risque de mauvaise planification du déploiement des antennes**.
* **Insatisfaction client** en cas de lenteur ou d'absence de signal.

### 1.5.2 Difficultés pour les utilisateurs finaux

* Qualité de service inégale
* Absence d'informations en cas de défaillance
* Aucune visibilité sur l'état du réseau en temps réel

## 1.6 Méthodes existantes d’analyse de données et de cartographie

### 1.6.1 Supervision classique

Utilisation de logiciels internes pour afficher des indicateurs, souvent sous forme de tableaux ou de cartes statiques. Peu de possibilités de détection spatiale automatique.

### 1.6.2 Systèmes d’Information Géographique (SIG)

Des outils comme QGIS permettent de visualiser les infrastructures, mais ils restent manuels et techniques, peu adaptés à l’analyse dynamique et à grande échelle.

## 1.7 Solution envisagée

Pour répondre à ces limites, nous proposons la conception d’une solution cartographique intelligente, capable de :

* Traiter a l’aide du machine learning des données raster géospatiales
* Identifier les zones densément peuplées ou mal desservies
* Regrouper ces zones selon leur profil technique
* Les afficher sur une carte interactive avec calques personnalisables

Cette solution vise à simplifier le travail des équipes techniques et à fournir une vue globale claire et actualisée du réseau, facilitant la planification, la maintenance, ou l’extension d’infrastructure.

### 1.7.1 Objectifs de la solution

* Automatiser l’identification des zones critiques
* Offrir une cartographie dynamique
* Permettre une lecture rapide et stratégique des données
* Renforcer l’aide à la décision dans les interventions et le déploiement

### 1.7.2 Description générale du système

Notre solution est composée de plusieurs étapes intégrées dans une pipeline simple et efficace :

* **Entrée** : des couches raster contenant plusieurs informations
* **Traitement automatique** : repérage des zones pertinentes, regroupement intelligent
* **Affichage** : carte interactive avec plusieurs calques, visualisable par n’importe quel agent

## ***Conclusion du chapitre***

Ce chapitre a présenté le contexte général dans lequel s’inscrit notre projet, en mettant en évidence les difficultés rencontrées par Djezzy dans l’analyse de ses zones d’activité, ainsi que les limites des approches actuelles. En explorant les méthodes existantes et en identifiant les besoins spécifiques de l’opérateur, nous avons défini une solution adaptée, innovante et réutilisable dans divers cas d’usage télécoms. Le chapitre suivant présentera cette solution en détail, en insistant sur son architecture, ses fonctionnalités et ses apports.

# 2. Chapitre 2 : Analyse et conception de la solution

Dans ce chapitre, nous détaillons la démarche adoptée pour concevoir et développer notre solution d’analyse géospatiale intelligente au sein de l’opérateur téléphonique Djezzy. L’objectif principal est de fournir un outil permettant d’extraire, segmenter et visualiser les zones critiques à partir de données raster complexes (densité de population, infrastructure, etc.), tout en garantissant des performances robustes et une grande lisibilité.

Le succès de ce projet repose sur deux piliers fondamentaux : la qualité des données d’entrée (format raster, géoréférencées) et la justesse de l’approche méthodologique pour le traitement spatial et la segmentation automatique. La solution met en synergie plusieurs outils et techniques issus de la data science, de la géomatique et du machine learning non supervisé.

## 2.1 Traitement des données raster

### 2.1.1 Importance cruciale de la qualité des données raster pour le traitement spatial

Dans tout projet basé sur l’analyse de données géospatiales, la qualité des fichiers sources joue un rôle déterminant. L’efficacité du pipeline d’analyse dépend directement de la précision, de la résolution et de la cohérence des couches raster utilisées. La moindre erreur dans les métadonnées géographiques, la projection spatiale ou les valeurs manquantes peut entraîner des décalages, des résultats incorrects, ou un échec complet de certaines étapes du traitement.

Dans le cadre de notre projet — une solution d’analyse automatique et de visualisation intelligente destinée à un opérateur télécom comme Djezzy — cet enjeu est encore plus crucial. En effet, la segmentation géographique d’un territoire, et l’interprétation des zones d’intérêt, ne peuvent être pertinentes que si les données d’entrée sont fiables et bien alignées.

Cette exigence est d’autant plus importante que :

* Les fichiers raster doivent être spatialement compatibles (même coordonnées, même résolution).
* Les valeurs “NoData” doivent être correctement gérées pour éviter les biais.
* La qualité des données conditionne directement la validité des clusters générés et leur lisibilité cartographique.

Les risques principaux liés à des données raster de mauvaise qualité incluent :

* Des regroupements erronés (zones mal alignées ou masques incorrects) ;
* Une mauvaise interprétation des indicateurs (densité, occupation du sol) ;
* Des visualisations illisibles ou imprécises ;
* Un traitement trop long, voire bloqué, en cas de données bruitées ou corrompues.

### 2.1.2 Sources et nature des données

Les données utilisées dans ce projet proviennent de deux sources ouvertes, fiables et régulièrement mises à jour :

* **Copernicus Human Settlement Layer** (<https://human-settlement.emergency.copernicus.eu/>) : cette plateforme fournit des couches raster à haute résolution représentant la densité de population et l’occupation du sol. Elle est reconnue à l’échelle internationale pour la qualité de ses produits géospatiaux dérivés d’images satellites.
* **OpenStreetMap (OSM)** : utilisé pour extraire les données vectorielles relatives aux bâtiments, routes et infrastructures. Ces données ont servi à générer une couche raster estimant la densité de constructions par unité de surface.

Afin d’adapter ces jeux de données à notre zone d’étude — la wilaya d’Alger — nous avons utilisé le logiciel **QGIS** pour :

* Effectuer un **découpage spatial** sur l’emprise géographique d’Alger ;
* Convertir les données vectorielles en raster lorsque nécessaire ;
* Harmoniser les systèmes de coordonnées (CRS) entre toutes les couches.

L'utilisation combinée de ces deux sources, couplée à une préparation minutieuse dans QGIS, a permis de disposer d’un jeu de données cohérent, ciblé et directement exploitable par notre pipeline de traitement.

## 2.2 Pipeline d'analyse et de segmentation

Cette section décrit les différentes étapes d'analyse et de traitement des données raster dans le but de produire une carte segmentée et interprétable. Le pipeline repose sur une combinaison de techniques de clustering non supervisé appliquées sur des données géospatiales extraites de couches raster. Il se déroule en trois grandes étapes : un filtrage initial pour extraire les zones denses, un regroupement intelligent par clustering, et une reconstruction spatiale pour affichage.

### 2.2.1 Filtrage spatial avec HDBSCAN

Le filtrage spatial est la première étape critique du pipeline. Son objectif est d'isoler les régions présentant une forte densité de données pertinentes (ex : forte population et concentration de bâtiments). Pour cela, nous utilisons l'algorithme **HDBSCAN** (Hierarchical Density-Based Spatial Clustering of Applications with Noise), qui est particulièrement adapté aux données bruitées et aux structures de densité inégale.

Contrairement à KMeans, HDBSCAN ne nécessite pas de spécifier un nombre de clusters à l'avance. Il se base sur la densité locale pour identifier les zones homogènes et supprime automatiquement les points trop isolés ou peu significatifs. Cela nous permet d'obtenir un **masque spatial** représentant uniquement les zones d'intérêt, tout en réduisant drastiquement le bruit dans les données.

Les avantages principaux de HDBSCAN dans notre contexte sont :

* La détection automatique du nombre de clusters potentiels (non supervisée)
* L'élimination naturelle des pixels dispersés (bruit)
* L'adaptabilité à la densité variable des données (zones urbaines vs périurbaines)

Le masque résultant est une première carte binaire où seuls les pixels appartenant à une zone dense sont conservés. Ce sous-ensemble devient la base du clustering suivant.

### 2.2.2 Regroupement par clustering KMeans (MiniBatch)

Après avoir isolé les zones pertinentes, nous appliquons un second niveau de regroupement : le clustering par **MiniBatchKMeans**. Il s'agit d'une variante de l'algorithme KMeans standard, plus rapide et plus adaptée à de grands ensembles de données.

L'objectif ici est de regrouper les pixels filtrés par similarité de caractéristiques (densité population, densité construction) afin d'obtenir des **zones cohérentes et homogènes**. Ces groupes sont appelés "clusters" et peuvent représenter différentes catégories de territoire : zones résidentielles denses, zones industrielles, quartiers périurbains, etc.

Le choix du nombre de clusters **K** est déterminé de manière empirique grâce à la **méthode du coude**. Cette méthode consiste à tracer l'inertie intra-classe (somme des distances des points à leur centroïde) en fonction du nombre de clusters. Le point de flexion sur cette courbe indique un bon compromis entre compacité des groupes et complexité du modèle.

Cette étape produit une étiquette de cluster pour chaque pixel filtré.

### 2.2.3 Attribution des labels et reconstruction géographique

Une fois les clusters générés, chaque pixel ayant survécu au filtrage HDBSCAN est associé à un label numérique (0, 1, 2, ...). Ces labels sont ensuite réinjectés dans une matrice 2D de mêmes dimensions que le raster d'origine (population ou construction).

Tous les pixels filtrés reçoivent un identifiant de cluster ; les autres sont marqués comme -1 ou laissés vides. Cela permet de **reconstruire une carte thématique**, où chaque région colorée correspond à un cluster donné.

Cette matrice segmentée est ensuite convertie en image RGB et superposée à une carte de base (via Folium) pour permettre une lecture géographique claire des résultats.

## 2.3 Visualisation cartographique

### 2.3.1 Choix de l'outil : Folium

Folium a été choisi pour sa compatibilité avec Leaflet.js et sa capacité à afficher :

* Des images raster (calques segmentés)
* Des calques vectoriels (GeoJSON)
* Des légendes personnalisées (via Branca)

### 2.3.2 Structure de la carte finale

La carte finale présente plusieurs calques interactifs :

* Un calque des clusters colorés
* Un calque optionnel des zones filtrées (HDBSCAN)
* Une légende dynamique par cluster (moyenne population, densité...)

### 2.3.3 Export et intégration

La carte est exportable en HTML, prêt à être intégrée dans un site web interne pour visualisation par les décideurs de Djezzy.

## 2.4 Architecture globale du pipeline

L’architecture générale du système se présente comme suit :

1. Import des rasters
2. Nettoyage / Extraction des pixels valides
3. Filtrage spatial (HDBSCAN)
4. Clustering KMeans (sur zones filtrées)
5. Reconstruction du masque 2D
6. Affichage sur carte interactive (Folium)

## 

Schéma du pipeline de traitement et de visualisation géospatiale

## 2.5 Contraintes techniques rencontrées

* Gérer les différents CRS et tailles raster
* Le masquage spatial (HDBSCAN → reconstruction propre)
* Détection automatique du bon nombre de clusters (stabilité de la méthode du coude)
* Lisibilité de la carte et poids à l’export (HTML)

## Conclusion du chapitre

Ce chapitre a exposé en détail notre démarche de conception : du traitement préalable des données raster jusqu’à l’affichage cartographique interactif. La stratégie adoptée nous a permis de concevoir un système souple, modulaire, et généralisable, répondant aux besoins de Djezzy en matière de visualisation intelligente des zones d’intérêt. Le prochain chapitre portera sur la mise en œuvre concrète, les tests réalisés, et l’analyse des résultats obtenus.