MSE | MASTER OF SCIENCE IN ENGINEERING

Hes·so
Haute Ecole Spécialisée de Suisse occidentale
Fachhochschule Westschweiz
University of Applied Sciences and Arts Western Switzerland

Master of Science HES-SO in Engineering
Av. de Provence 6
CH-1007 Lausanne

Master of Science HES-SO in Engineering

Orientation : Technologies de l'information et de la communication (TIC)

# AI-enhanced LoRa Based Indoor Localization System

Fait par
Manon Racine

Sous la direction de
Dr. Nuria Pazos
HE-Arc

St-Imier, HE-Arc, 19 octobre 2018

# Table des matières

**1 Rapport intérmédiaire : 17.09.2018 au 12.10.2018** **1**
1.1 Cahier des charges . . . 1
1.1.1 Introduction . . . 1
1.1.2 But du projet . . . 1
1.1.3 Objectifs et tâches à réaliser . . . 1
1.2 Résumé du document 00 . . . 2
1.3 Résumé du document 01 . . . 2
1.3.1 Introduction . . . 2
1.3.2 Algorithmes . . . 3
1.3.3 Conclusions . . . 5
1.4 Résumé du document 02 . . . 8
1.4.1 Ranging Errors . . . 9
1.4.2 Algrorithme . . . 9
1.4.3 Conclusion . . . 9
1.5 Résumé du document 03 . . . 10
1.5.1 A comparative review of fingerprinting's performance . . . 11
1.5.2 A confidence machine approach to fingerprinting . . . 12
1.5.3 Test bed . . . 12
1.5.4 Conclusion . . . 13
1.6 Choix de l'algorithme . . . 14

**2 Rapport intérmédiaire : 15.10.2018 au 26.10.2018** **15**
2.1 Objectifs et tâches à réaliser . . . 15

**3 Introduction** **17**
3.1 Contexte . . . 17
3.2 Aspect Novateur . . . 18
3.3 Objectifs et tâches à réaliser . . . 18
3.4 Structure du rapport . . . 18

**4 Etat de l'art** **19**
4.1 Extrait de choses importantes . . . 19
4.1.1 Type d'apprentissage . . . 19
4.1.2 Algorithme . . . 19
4.1.3 Outliers [FP17] . . . 21
4.2 Comparaison . . . 21
4.3 Choix . . . 21

**Bibliographie** **23**

# Table des figures

1.1 The new attribute “cell” construction phase . . . 3
1.2 Accuracy results of classification using whole dataset . . . 6
1.3 Accuracy results of building classification . . . 6
1.4 Accuracy results of floor classification . . . 7
1.5 Accuracy results of region classification . . . 7
1.6 Accuracy results of machine learning algorithms . . . 8
1.7 Elapsed time results of machine learning algorithms . . . 8
1.8 Ranging error - LOS and NLOS . . . 9
1.9 Les deux phases du fingerprinting . . . 10
1.10 Performance accuracy of fingerprinting at Microsoft IPSN 2014 competition . . . 11
1.11 Performance accuracy of W-KNN, Naïve Bayes and neural network, reported in first review 12
1.12 Summary of the three fingerprinting test beds used in this article . . . 13
1.13 Comment choisir un algorithme ML . . . 14

3.1 Etat de l’art des différentes méthode de positionnement intérieur [LLI] . . . 17

# 1 Rapport intérmédiaire : 17.09.2018 au 12.10.2018

Ce premier résumé a pour but de poser le projet et d'étudier l'état de l'art. Depuis les lectures concernant ce qu'il existe en machin learning pour le positionnement indoor, il est nécessaire de faire un résumé afin de choisir le meilleur algorithme afin d'améliorer le positionnement indoor à l'aide de la technologie LoRa et le mode ranging.

## 1.1 Cahier des charges

### 1.1.1 Introduction

Les systèmes de localisation basés sur GPS souffrent de la détérioration de la précision et sont presque indisponibles dans les environnements intérieurs. Pour les environnements intérieurs, de nombreuses technologies de systèmes de positionnement ont été conçues sur la base de la vision, de la détection infrarouge ou ultrasonore, des champs magnétiques de la terre, des accéléromètres / gyromètres, des balises BLE ou de la communication WiFi. Chacune de ces technologies existantes a des coûts, une précision et un compromis maximum en matière de couverture, mais un service de localisation intérieur générique reste difficile à obtenir.

### 1.1.2 But du projet

S'appuyant sur les capacités étendues des nouveaux circuits intégrés LoRa, ce projet développera et déploiera un système de localisation capable d'améliorer la précision de la position atteinte par les systèmes de localisation basés sur LoRa existants reposant sur des mécanismes TDOA ou de télémétrie. À cette fin, une exploration et une comparaison des différentes techniques "machine learning/deep learning" pour le positionnement basé sur le "fingerprinting" seront effectuées.

### 1.1.3 Objectifs et tâches à réaliser

1. Etudier le cahier des charges
2. Etudier l'état de l'art des techniques à utiliser dans le cadre du projet, en particulier les systèmes de localisation indoor basés sur des techniques d'apprentissage, et réunir une documentation (env. 20
3. Etablir un planning pour l'ensemble du projet.
4. Définir un plan des tests à effectuer.
5. Définir les procédures de test
6. Définir le setup pour la collecte de données de localisation
7. Prise en main de l'environnement de développement pour les phases de training et du test de la technique d'apprentissage retenue (e.g., PyTorch).
8. Implémentation de la solution ML retenue.
9. Tester le système selon le protocole préétabli.
10. Faire des propositions pour améliorer les performances de l'algorithme et, si possible, les implémenter.
11. Rédiger le rapport et documenter l'ensemble du projet.

## 1.2 Résumé du document 00

Ce document décrit la localisation sans utiliser le GPS et en utilisant LoRa. Il a surtout été utile d'utiliser ce document pour la gestion des "outliers"

Le problème principal avec le GPS c'est la consommation et la durée de vie c'est pourquoi un système basé sur LoRa a été étudié. La portée en milieu rural est d'environ 15km alors qu'il est de 5km dans un milieu urbain cela grâce à la bonne sensibilité du récepteur (-130dBm). Une chose intéressant est la bande passante qui est plus large que d'autres technologies qui permet de distinguer différents chemins du même signal. Sagemcom ont obtenus des bons résultats au niveau de la précision qui est de environ 4 mètres.

Ce qui est intéressant c'est dans cette publication c'est la manière de traiter les "outliers - valeurs aberrante, c'est-à-dire les points qui ne sont pas cohérent lors d'une mesure. Selon Barnet et Lewis [11], un "outliers" est définit comme étant une observation qui semble incompatible avec le reste d'un ensemble de données. Garder un "outliers" dans une set de données peut amener à de mauvais résultat il est donc important de les détecter correctement. Il existe différentes méthodes pour déterminer ces "outliers" :

1. Grubbs' test : Détecte un "outliers" en supposant une distribution normale.
2. Tietjen-Moore test : C'est une généralisation de Grubbs' test pour détecter de multiple outliers. Il a cependant un inconvénient, il est nécessaire de connaitre le nombre exact d'ouliers.
3. Generalized Extreme Studentized Deviate (ESD) : C'est également une généralisation du test Grubbs' mais il n'est pas nécessaire de connaitre à l'avance le nombre d'ouliers. Ce test nécessite uniquement une limite supérieure pour le nombre suspect d'outliers.[01]

[11] V. Barnett ; T. Lewis, Outliers in Statistical Data, 3rd ed. Wiley Series in Probability and Mathematical Statistics, 1994.

[01] lien concernant les ESD : https ://www.itl.nist.gov/div898/handbook/eda/section3/eda35h3.htm

## 1.3 Résumé du document 01

Cette publication parle d'une analyse comparative entre différents algorithmes de "machine learning" pour du positionnement indoor. L'étude est basée sur un positionnement "fingerprint" ce qui permet de cartographier un endroit à l'aide du de la force du signal récéptionné (RSS - Received Signal Strength). Dans cet article, les algorithmes de "machine learning" sélectionnés sont comparés en termes de précision de positionnement et de temps de calcul. la base de donnée UJIIndoorLoc a été utilisé pour les différentes expérimentation. Les résultats expérimentaux révèlent que l'algorithme k-Nearest Neighbor (k-NN) est le plus approprié lors du positionnement.

### 1.3.1 Introduction

Au cours des expériences, la base de données UJIIndoorLoc, qui est préparée pour les systèmes de positionnement à l'intérieur [8], est utilisée. La classification est effectuée en premier lieu en utilisant le jeu de données d'origine en considérant les valeurs RSS de 520 points d'accès sans fil (WAP) et les nouveaux attributs définis en tant que «cellule» qui composent les attributs BuildingID, Floor, SpaceID et RelativePos. Ensuite, une nouvelle méthode est proposée : «Séparation déductive pour le positionnement intérieur (DESIP - Deductive Separation for Indoor Positioning)». Dans cette méthode, tout d'abord, seules les informations de bâtiment et les valeurs RSS mesurées à partir de 520 WAP sont utilisées pour la tâche de classification.

Durant les expériences, des algorithmes déterministes tels que le plus proche voisin (NN - nearest neighbor), le SMO, l'arbre de décision (J48) et des algorithmes probabilistes tels que Naïve Bayes et Bayes Net sont utilisés. L'algorithme le plus approprié pour la solution du problème de positionnement intérieur est déterminé en comparant la précision et le temps de calcul de chaque approche.

La base de données entière est séparée de telle sorte que 19.937 enregistrements soient réservés à la formation et 1.111 enregistrements soient réservés aux tests. Il y a 529 caractéristiques et ces caractéristiques sont les coordonnées où sont prises les empreintes digitales WiFi, telles que bâtiment, étage, espace (bureau, laboratoire, etc.), position relative (dans une pièce ou dans un couloir), etc. Le jeu de données de formation UJIIndoorLoc comprenant les valeurs RSS de 520 WAP et un nouvel attribut «cellule» qui compose les attributs floor, buildingID, spaceID et relative position de l'ensemble de données d'origine est utilisé pour la tâche de classification. Les étapes des expériences utilisant ce jeu de données sont illustrées à la figure 1.1.

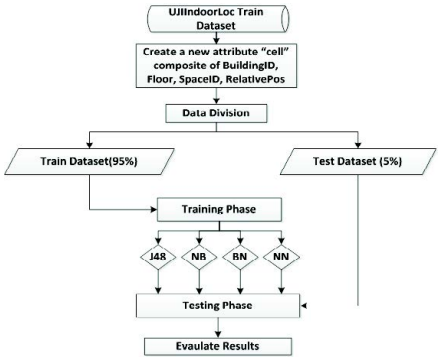


**Figure 1.1 –** The new attribute "cell" construction phase

### 1.3.2 Algorithmes

Dans la section suivante, les algorithmes de classification utilisés dans cette étude sont brièvement décrits.

#### Decision Tree

L'arbre de décisions et une méthode très connue en "machine learning". Il possède des noeuds de décisions (non-terminal), des branches, et des noeuds feuilles (terminal) qui représentent les caractéristiques, condition et les classes. A chaque noeud de décision on sait quelle branches suivre et lorsque l'algorithme atteint un noeud final, le label contenu dans ce même noeud est retourné comme étant la classe. L'ID3 de Quinlan et son successeur, C4.5, sont les plus populaires parmi les algorithmes d'arbre de décision [19].

(19) J. R. Quinlan, " C4. 5 : programs for machine learning", Elsevier, 2014.

#### Naïve Bayes

Le classificateur Naïve Bayes [22] basé sur le théorème de Bayes est un algorithme d'apprentissage supervisé [23]. Il est robuste aux données bruyantes, facile à construire, affiche une grande précision et rapidité lorsqu'il est appliqué à de grandes bases de données et exécute des modèles de classification plus complexes. Par conséquent, il est largement utilisé dans les tâches de classification. Il calcule la probabilité de chaque attribut dans les données en supposant qu'elles sont également importantes et indépendantes les unes des autres. Cette hypothèse est appelée indépendance conditionnelle de classe [24, 25].

(22) G. H. John, and P. Langley, "Estimating Continuous Distributions in Bayesian Classifiers", 11th Conference on Uncertainty in Artificial Intelligence, pp., 338-345, 1995.

(23) C. Anuradha, and S. Dhall, "Software Defect Prediction Using Supervised Learning Algorithm and Unsupervised Learning Algorithm", 2013.

(24) W. Yotsawat, and A. Srivihok, "Inbound tourists segmentation with combined algorithms using K-Means and Decision Tree", 10th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp.189-194, 2013.

(25) S. Ureerat, and P. Singsri, "The classifier model for prediction quail gender after birth based on external factors of quail egg", IEEE 11th International Joint Conference on Computer Science and Software Engineering (JCSSE), 2014.

### Bayesian Network

L'algorithme de réseau bayésien est largement utilisé pour la classification et est basé sur le théorème de Bayes où la probabilité conditionnelle sur chaque nœud est calculée et forme un réseau bayésien. Il s'appelle également réseau de croyance ou réseau occasionnel. Réseau bayésien a deux parties nommées qualitatives et quantitatives, qui sont la structure topologique du réseau bayésien et le tableau de probabilité conditionnelle (CPT), respectivement [26].

Le réseau bayésien est un graphe acyclique dirigé où chaque nœud représente un attribut des données et un ensemble de distributions de probabilité. Ces distributions donnent les probabilités pour la valeur de chaque nœud étant donné que les parents de nœud.

(26) D. Yang, and L. Jin-lin, "Research on personal credit evaluation model based on bayesian network and association rules", 2007 International Conference on Wireless Communications, Networking and Mobile Computing, 2007.

### K-Nearest Neighbor

Le classificateur K-Nearest Neighbor (K-NN) [27] est également connu sous le nom de classificateur basé sur la distance qui classe les instances en fonction de leur similarité. C'est l'un des algorithmes les plus populaires de l'apprentissage automatique. C'est un type d'apprentissage paresseux dans lequel la fonction n'est approchée que localement et tout calcul est retardé jusqu'à la classification. Le tuple inconnu dans K-NN est assigné à la classe la plus commune parmi ses K-plus proches voisins. Lorsque K = 1, le tuple inconnu se voit attribuer la classe du tuple d'apprentissage le plus proche dans l'espace des motifs [28].

(27) D. W. Aha, D. Kibler , and M. K. Albert, " Instance-based learning algorithms", Machine Learning, vol. 6, pp., 37-66, 1991.

(28) C. Shah, and A. G. Jivani, "Comparison of data mining classification algorithms for breast cancer prediction", 2013 Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT), pp.1-4, 2013.

### SMO

L'algorithme d'optimisation séquentielle minimale (SMO - Sequential minimal optimization) [29] est représenté par John C. Platt pour la formation du classificateur de vecteurs de support à l'aide des noyaux polynomiaux ou RBF. C'est l'un des algorithmes les plus courants pour la classification des grandes marges par SVM. Il remplace globalement toutes les valeurs manquantes et transforme les attributs nominaux en attributs binaires. La SVM est une technique de classification basée sur la technologie des réseaux neuronaux utilisant la théorie de l'apprentissage statistique [30]. Il recherche un hyperplan optimal linéaire afin de maximiser la marge de séparation entre la classe positive et la classe négative. En pratique, la plupart des données ne sont pas linéairement séparables ; ainsi, pour rendre la séparation possible, la transformation est effectuée à l'aide d'une fonction du noyau. L'entrée est transformée en un espace caractéristique de dimension supérieure à l'aide d'une cartographie non linéaire [30]. Une décision sur la fonction du Kernel est nécessaire pour implémenter SVM. Le Kernel définit la classe de fonction [31].

[29] J. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization", Advances in Kernel Methods - Support Vector Learning, 1998.

[30] P. Niken, and H. Ohwada, "Applicability of machine-learning techniques in predicting customer defection", IEEE 2014 International Symposium on Technology Management and Emerging Technologies (ISTMET), 2014.

[31] S. M. Obaidullah, K. Roy, and N. Das, "Comparison of different classifiers for script identification from handwritten document", 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), pp.1-6, 2013.

#### AdaBoost

AdaBoost (Adaptive Boosting) [32] est un algorithme d'apprentissage d'ensemble. Généralement, il peut être utilisé avec des algorithmes de Machine learning faibles pour améliorer leurs performances. Il est simple à mettre en œuvre, rapide et moins susceptible d'avoir un overfitting. Il améliore les algorithmes de classification instables tels que J48, DecisionStump, etc. L'idée derrière cet algorithme est d'obtenir un classificateur très précis en combinant de nombreux classificateurs faibles. Il fonctionne en exécutant de manière répétée un algorithme d'apprentissage faible donné sur diverses distributions sur les données d'apprentissage, puis en combinant les classificateurs produits par l'apprenant faible en un classificateur composite unique [33]. Les classificateurs de l'ensemble sont ajoutés un par un, de sorte que chaque classificateur suivant est entrainé sur des données difficiles pour les membres précédents de l'ensemble. Les poids sont définis sur les instances du jeu de données, en suivant une règle selon laquelle les instances difficiles à classer prennent plus de poids. Cette règle conduit les classificateurs ultérieurs à se concentrer sur eux [34].

[32] Y. Freund, and R. E. Schapire, "Experiments with a new boosting algorithm", 3th International Conference on Machine Learning, San Francisco, pp. 148-156, 1996.

[33] R. Shams, and R. E. Mercer, "Classifying Spam Emails Using Text and Readability Features", 2013 IEEE 13th International Conference on Data Mining (ICDM, pp. 657-666, 2013.

[34] S. O. Sharif, L. I. Kuncheva, and S. P. Mansoor, "Classifying encryption algorithms using pattern recognition techniques", 2010 IEEE International Conference on Information Theory and Information Security (ICITIS), , pp. 1168-1172, 2010.

#### Bagging

le Bagging [35] crée des sacs de données de la même taille que le jeu de données d'origine en appliquant une sélection aléatoire à différents sous-ensembles des données d'apprentissage avec de nombreux exemples qui apparaissent plusieurs fois. Ce processus est appelé réplication bootstrap des données d'entrainement. L'idée derrière cette technique est de construire différents classificateurs en utilisant ces sous-ensembles. Chaque sous-ensemble est utilisé pour entrainer un classificateur individuel. Cette approche d'ensemble utilise le nombre de classificateurs a priori [35].

[35] L. Breiman, "Bagging predictors", Machine Learning. vol. 24, no. 2, pp.123-140, 1996.

### 1.3.3 Conclusions

Dans cet article les algorithme suivant ont été comparé : NN, SMO, J48, Naïve Bayes and BayesNet.

Lorsque tout le dataset est pris en compte, c'est l'algorithme J48 qui est le meilleur :

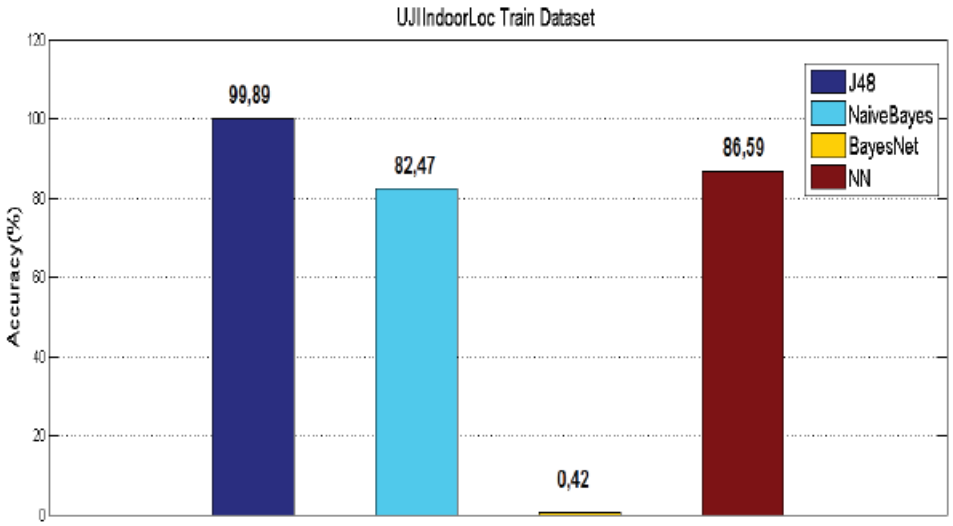


**Figure 1.2 –** Accuracy results of classification using whole dataset

Ensuite, en accordance à l'approche DESIP (Deductive Separation for Indoor Positioning), la classification est effectuée en 3 phases (building,floor and region). Le resultat des algorithme pour cette classification donne BayesNet comme étant le meilleur.

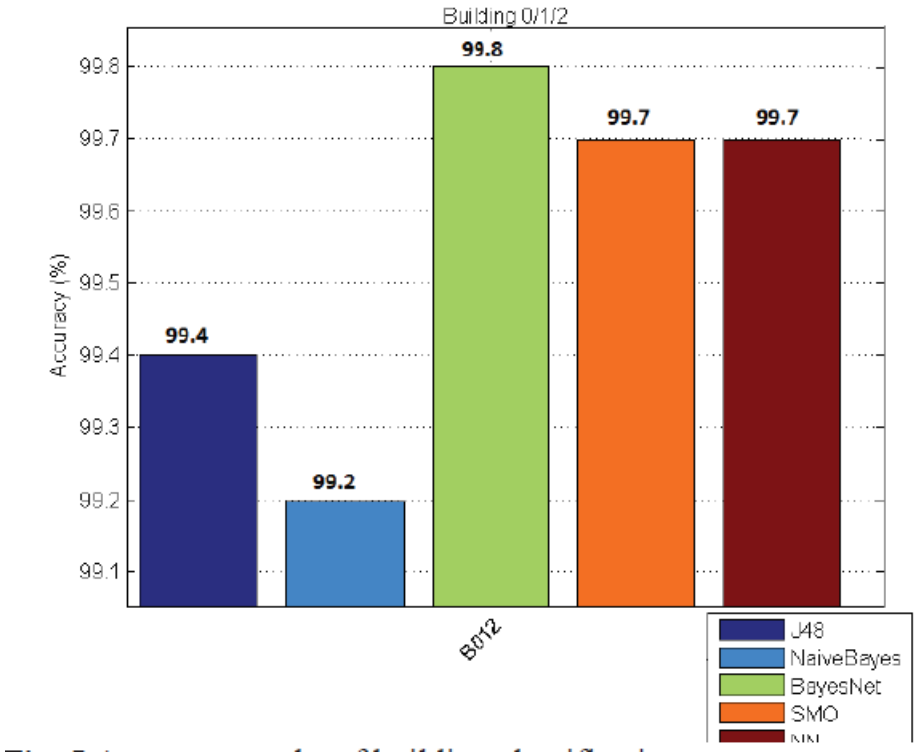


**Figure 1.3 –** Accuracy results of building classification

Suite à cela la classification a été faite en fonction des étages (floors). Dans ce cas le meilleur algorithme est NN.

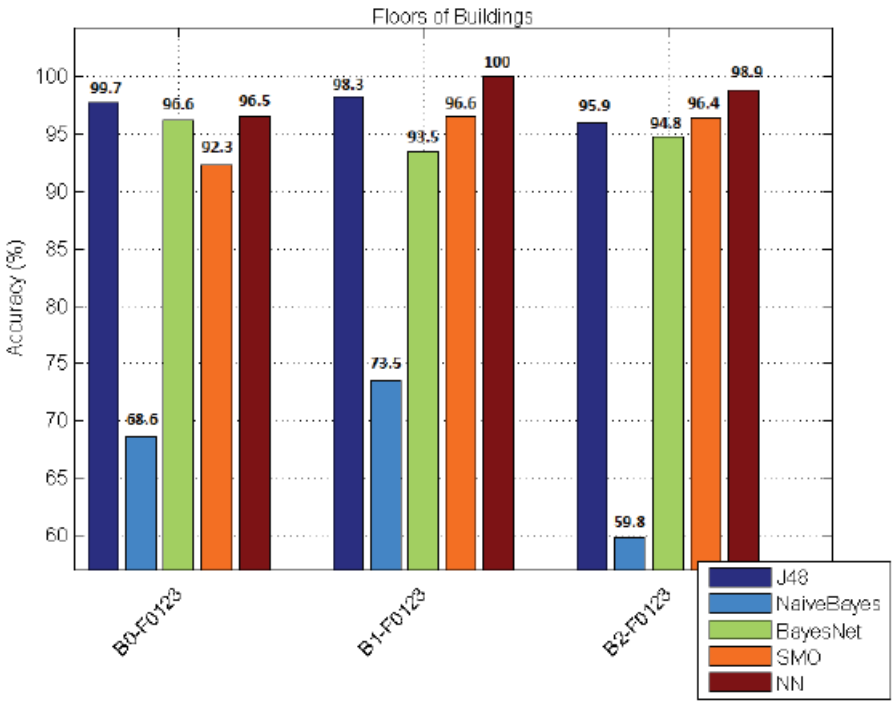


**Figure 1.4 –** Accuracy results of floor classification

Et pour la dernière étape, la classification a été faite en fonction de la région et là encore c'est l'algorithme NN qui est le meilleur.

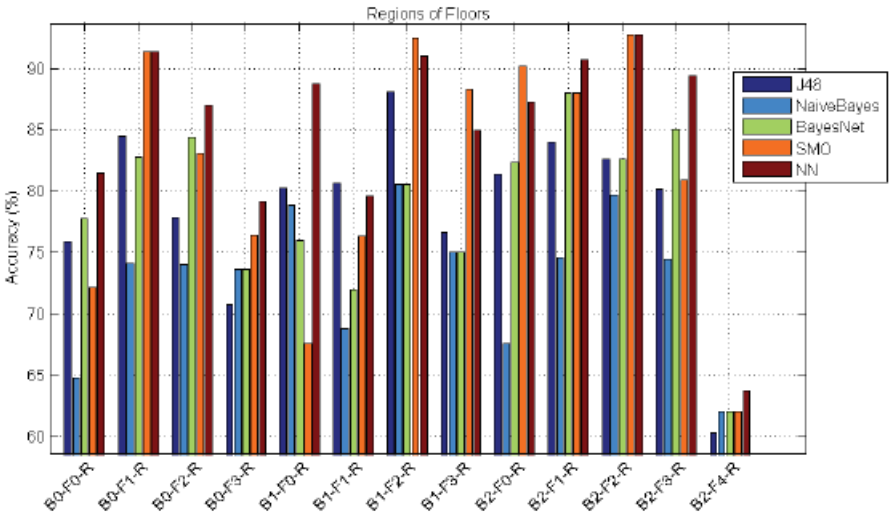


**Figure 1.5 –** Accuracy results of region classification

Si les deux tableaux ci-dessous sont analysés, l'algorithme NN est le meilleurs pour tous les dataset niveau temps d'execution. En ce qui concerne la précision, Bayes Net est meilleur pour la classification "building" par contre NN est meilleur dans tous les autres cas.

| | J48 | Naive Bayes | Bayes Net | SMO | NN |
|---|---|---|---|---|---|
| B012 | 99,4 | 99,19 | 99,80 | 99,69 | 99,69 |
| B0_F0123 | 97,7 | 68,58 | 96,17 | 92,34 | 96,55 |
| B0_F0_R | 75,93 | 64,81 | 77,78 | 72,22 | 81,48 |
| B0_F1_R | 84,48 | 74,14 | 82,76 | 91,37 | 91,38 |
| B0_F2_R | 77,92 | 74,03 | 84,42 | 83,12 | 87,01 |
| B0_F3_R | 70,83 | 73,61 | 73,61 | 76,39 | 79,17 |
| B1_F0123 | 98,28 | 73,54 | 93,47 | 96,56 | 100 |
| B1_F0_R | 80,28 | 78,87 | 76,06 | 67,61 | 88,73 |
| B1_F1_R | 80,64 | 68,82 | 72,04 | 76,34 | 79,57 |
| B1_F2_R | 88,06 | 80,60 | 80,59 | 92,54 | 91,04 |
| B1_F3_R | 76,67 | 75 | 75 | 88,33 | 85 |
| B2_F01234 | 95,95 | 59,77 | 94,83 | 96,40 | 98,87 |
| B2_F0_R | 81,37 | 67,65 | 82,35 | 90,19 | 87,25 |
| B2_F1_R | 84 | 74,67 | 88 | 88 | 90,67 |
| B2_F2_R | 82,61 | 79,71 | 82,61 | 92,75 | 92,75 |

**Figure 1.6 –** Accuracy results of machine learning algorithms

| | J48 | Naive Bayes | Bayes Net | SMO | NN |
|---|---|---|---|---|---|
| B012 | 62,41 | 2,5 | 11,19 | 7,25 | 0,06 |
| B0_F0123 | 11,74 | 0,65 | 2,03 | 14,31 | 0,01 |
| B0_F0_R | 1,5 | 0,22 | 0,7 | 14,1 | 0 |
| B0_F1_R | 1,71 | 0,21 | 0,88 | 17,93 | 0 |
| B0_F2_R | 2,52 | 0,2 | 0,9 | 18,77 | 0 |
| B0_F3_R | 1,62 | 0,2 | 0,88 | 17,3 | 0 |
| B1_F0123 | 10,05 | 0,57 | 2,13 | 11,5 | 0,02 |
| B1_F0_R | 1,68 | 0,18 | 0,87 | 18,22 | 0 |
| B1_F1_R | 1,75 | 0,21 | 1,01 | 13,3 | 0 |
| B1_F2_R | 1,96 | 0,27 | 1,11 | 19,47 | 0 |
| B1_F3_R | 0,95 | 0,13 | 1,11 | 12,1 | 0 |
| B2_F01234 | 18,09 | 1,18 | 4,4 | 18,81 | 0,03 |
| B2_F0_R | 2,54 | 0,24 | 1,11 | 21,57 | 0 |
| B2_F1_R | 3,27 | 0,29 | 1,15 | 27,1 | 0 |
| B2_F2_R | 1,86 | 0,2 | 1,46 | 19,76 | 0 |

**Figure 1.7 –** Elapsed time results of machine learning algorithms

De cette publication en découle que NN est supérieur à toutes les autres méthodes pour estimer la position. En outre, J48 offre des performances quasiment identiques lorsqu'il est utilisé avec des algorithmes itératifs, à savoir AdaBoost et Bagging.

## 1.4 Résumé du document 02

Cette publication traite de la diminution de erreur du mode ranging concernant la localisation UWB (ultra wide band). Plusieurs techniques existent pour diminuer l'erreur de positionnement en détectant ce qui est en ligne de vue (LOS) ou non (NLOS). Ici, il est exploité une autre technique qui va directement diminuer cet erreur que ça soit en LOS ou NLOS. Ils appliquent deux classes de régresseurs non paramétriques pour avoir une estimation de l'erreur de mesure. Afin de valider leurs résultats ils ont fait un vaste campagne de mesures intérieures. Cette technique montre une amélioration de performances significatives dans divers scénarios par rapport aux approches conventionnelles.

Ils se sont appuyé sur des outils de Machine Learning, et proposent deux techniques de régression non paramétriques pour estimer l'erreur de mesure, en se basant uniquement sur la forme d'onde reçue et la distance estimée.

1. La première technique utilise une régression de machine à vecteurs de support (SVM - support vector machine) pour trouver un hyperplan qui se rapproche de l'erreur de mesure en fonction des données d'apprentissage.
2. La seconde technique utilise un processus gaussien (GP) pour déterminer la distribution a posteriori de l'erreur de mesure, en fonction des données d'apprentissage.

L'erreur de mesure estimée, associée à une mesure de certitude, peut être transmise à un algorithme de localisation. Leurs techniques de régression présentent l'avantage supplémentaire de pouvoir être appliquées même lorsque les données d'apprentissage ne sont pas étiquetées avec des informations LOS ou NLOS.

### 1.4.1 Ranging Errors

Lors de mesure il existe beaucoup de paramètres qui peuvent créer une erreur. Avec un seul model il est difficile de capturer tout les types de perturbations. Dans cette publication, ils se basent sur 1024 mesures (512 LOS et 512 NLOS).

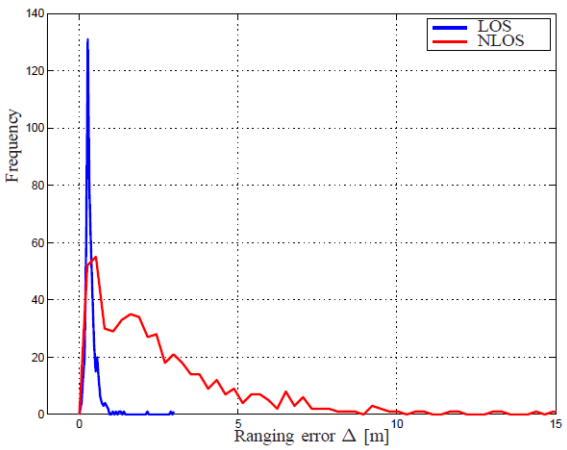


Figure 1. Histogram of the ranging error for the LOS and NLOS condition.

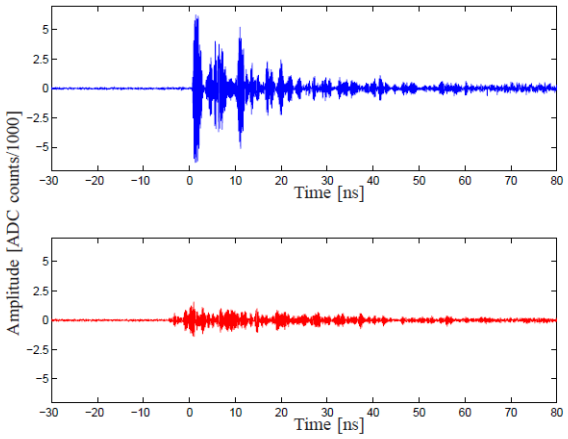


Figure 2. In some situations there is a clear difference between LOS (upper waveform) and NLOS (lower waveform) signals.

**Figure 1.8 –** Ranging error - LOS and NLOS

Suite, aux différentes mesures et différentes observations dans leur article, ils décident de ne pas labéliser les signaux en LOS et NLOS.

### 1.4.2 Algrorithme

#### Regression with Support Vector Machines

#### Regression with Gaussian Processes

### 1.4.3 Conclusion

Les approches classiques pour faire face au défi de la localisation dans des environnements encombrés impliquent généralement d'abord la détection de la condition NLOS, puis la prise de mesures appropriées pour prendre en compte la condition NLOS. Toutefois, la grande variété de matériaux et d'environnements

d'exploitation variés peuvent impacter les performances de la mesure de ranging, ce qui indique que la distinction entre LOS et NLOS n'est pas toujours significative. Sur la base de cette observation, ils ont adopté pour une approche différente dans cet article. Leur approche utilise des techniques de Machine Learning non paramétriques (SVM et GP) pour estimer l'erreur de "ranging" directement à partir de la forme d'onde reçue, sans aucune connaissance a priori ou a posteriori de la condition NLOS.

Sur la base d'une vaste campagne de mesures en intérieur avec des radios UWB conformes à la FCC, ils ont évalué les performances de localisation en termes de probabilité de panne pour différentes stratégies de localisation. Leurs résultats ont révélé que :

1. La minimisation de l1-norme est plus robuste pour faire face aux valeurs aberrantes (outliers) que la minimisation de l2-norme, pour une localisation sans atténuation.
2. les contraintes peuvent générer des gains significatifs, en particulier lorsque les exigences de localisation ne sont pas trop strictes.
3. les techniques de régression SVM ou GP offrent des gains de performance supplémentaires pour tous les scénarios considérés.
4. Les techniques de régression SVM ou GP, combinées à la connaissance des contraintes relatives à l'erreur de "ranging", offrent les meilleures performances pour les scénarios considérés.

## 1.5 Résumé du document 03

Pareil que dans les trois précédentes publications, le but est d'améliorer la précision du positionnement intérieur où il n'est pas possible d'utiliser un GPS. Cela toujours en tenant compte des problèmatiques d'un environnement dynamique avec des personnes qui bougent et de l'environnement complexe avec des murs, etc.. Dans ce papier ils ont validé leur solution dans 3 immeubles. Cet article est basé sur un positionnement WIFI fingerprinting et se base sur la force du signal recu.

Pour effectuer les mesures il y a deux phases.

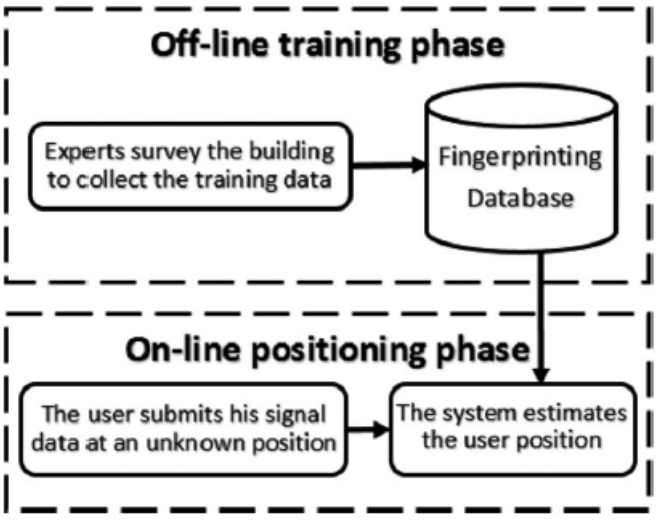


**Figure 1.9 –** Les deux phases du fingerprinting

La première consiste à décider combien de mesure on veut faire (tous les metres), comment sont prises les mesures (plusieurs mesures sont faites à chaque position) et comment labéliser le signal (souvent labélisé avec la coordonnée réelle). Cette phase est très importante et il est nécessaire de bien réfléchir comment procéder. (utiliser un robot par exemple)

Concernant la seconde phase c'est de définr le positionement. Durant cette phase la partie délicate est de définir quel algorithme utiliser.

### 1.5.1 A comparative review of fingerprinting's performance

Il existe une compétition comparative pour les positionnements indoor (Microsoft IPSN 2014).

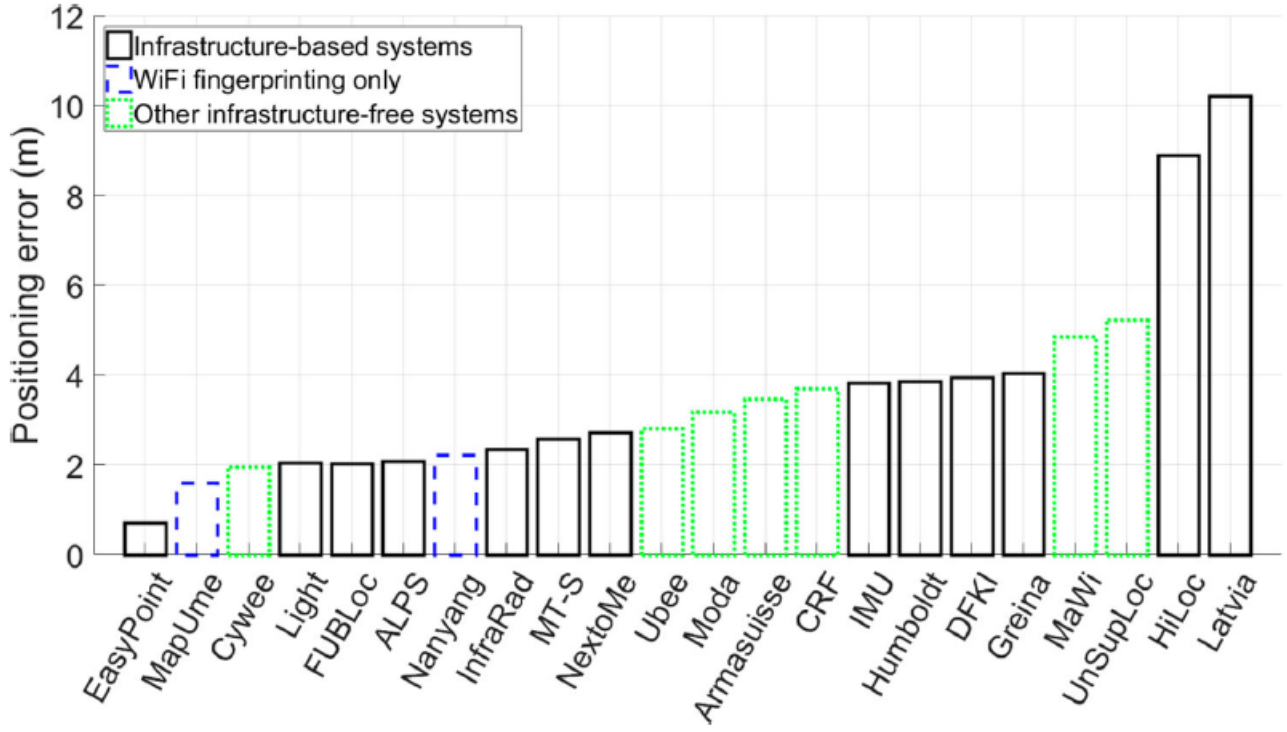


**Figure 1.10 –** Performance accuracy of fingerprinting at Microsoft IPSN 2014 competition

Cela permet de mettre en évidence les précisions qui sont obtenues. Les algortihmes qui seront étudiés sont : Weighted K-nearest neighbours (W-KNN), Naïve Bayes, neural network and histogram. Tous les mesures et tests sont bassé sur le RSS du WIFI.

Les mesures ont été effectuées de trois manières différentes (voir dans le document)

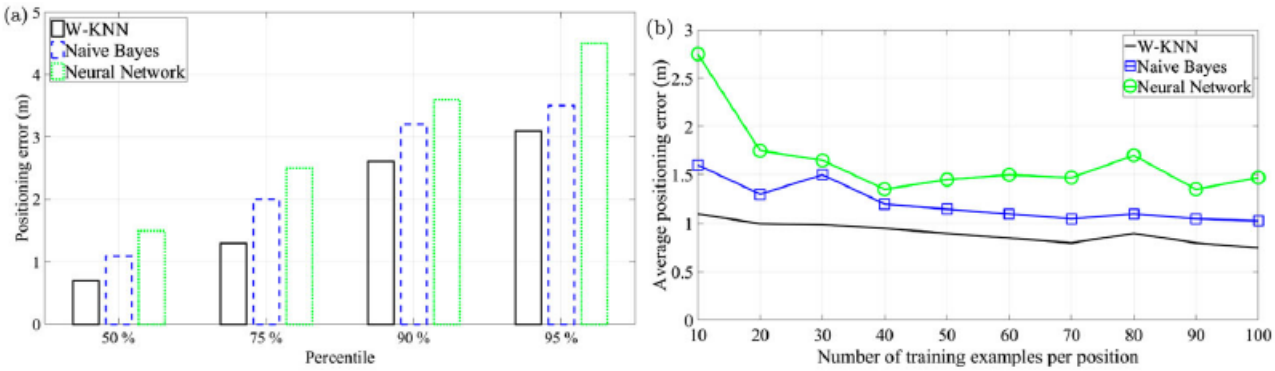


**Figure 3.** Performance accuracy of W-KNN, Naïve Bayes and neural network, reported in first review (Lin & Lin, 2005). (a) Performance accuracy of the three algorithms and (b) impact on the number of training examples.

**Figure 1.11 –** Performance accuracy of W-KNN, Naïve Bayes and neural network, reported in first review

En résumé, la systèse des trois solutions suggèrent que, avec uniquement la mesure métrique WiFi RSS, de nombreux algorithmes complexes risquent de ne pas être aussi performants que des algorithmes plus simples. Malgré sa simplicité, W-KNN a excellé dans la plupart des analyses de "fingerprinting". Il convient de noter que le système MapUme, deuxième sur 22 concurrents du concours Microsoft IPSN 2014, utilisait également W-KNN comme principal algorithme. Cependant, l'approche Naïve Bayes améliore sa précision lorsque le nombre d'entrainements est élevé, ce qui indique qu'au-delà du WiFi RSS, des informations supplémentaires seront nécessaires pour améliorer d'avantage les performances du "fingerprinting".

### 1.5.2 A confidence machine approach to fingerprinting

Une chose qui est égalemnt importante c'est la confiance qu'il y a dans une algorithme. Pour cela cet article à utilisé "conformal prediction (CP)" afin de donner un indice de confiance.

Plus on souhaite de confiance plus il est nécessaire d'avoir de set de données. Les avantages d'utiliser l'indice de confiance sont les suivants :

1. Chaque prédiction est associée à un indice de confiance et permet de dire combien la prédiction est correcte.
2. La prédiction produite par CP est statistiquement correcte sous les paramètres qui ont été choisi dans la phase on-line.
3. le niveau de confiance peut être ajusté pour produire un ensemble de prédiction plus grand ou plus petit.

Dans cet article, l'algorithme W-KNN a été utilisé.

### 1.5.3 Test bed

Afin de valider leur algorithme, ils ont utilisé 3 bancs de test.

**Table 4.** Summary of the three fingerprinting test beds used in this article.

| | Royal Holloway | Cambridge | UJIIndoorLoc |
|---|---|---|---|
| Abbreviation | RH | Cam | UJI |
| Training examples | 13,600 | 78,000 | 19,937 |
| Training area | 1480 $m^2$ | 540 $m^2$ | 110,000 $m^2$ |
| Surveyed space | 3 corridors | 1 corridor, 1 room | 3 buildings |
| Training time | 1 day | 2 days | 20 days |
| Granularity | 1 m | 10 cm for room<br>30 cm for corridor | Up to 2 m |
| Test samples | 150 | 100 | 1,111 |
| Last test sample | Same day | 1 day later | 3 months later |
| Measuring time | Working hours | Weekend | Working hours |
| Measures per location | 200 | 40 | Up to 30 |
| Distinct positions | 68 | 1950 | 933 |
| Orientations per location | 4 | 4 | Unknown |
| Label type | Cartesian<br>(metre scale) | Cartesian<br>(metre scale) | Longitude<br>and latitude |
| Label generator | Manually<br>by the surveyor | Automatically<br>by Active Bat system | Manually<br>by the surveyor |
| Total building(s) | 1 | 1 | 3 |
| Total floor(s) | 1 | 1 | 13 |
| WiFi APs | 131 | 43 | 529 |
| Device(s) used | 1 phone | 1 netbook | 25 phones |
| Surveyor(s) | 1 person | 1 robot | 18 people |

**Figure 1.12 –** Summary of the three fingerprinting test beds used in this article

1. Royal Holloway : Données récoltées manuellement dans un office standard avec un smartphone
2. Cambridge : Données récoltées automatiquement par un robot dans un environnement assez idéal.
3. UJIIndoorLoc : Utilise le dataset publique qui couvre une grande surface indoor basé sur trois bâtiments.

### 1.5.4 Conclusion

Cet article propose une nouvelle approche d'apprentissage basé sur la confiance d'un algorithme permettant d'estimer la position de l'utilisateur à l'intérieur avec la force du signal WiFi. Il introduit une mesure de confiance, non seulement utile pour refléter l'incertitude des prédictions de positionnement, mais également capable d'ajuster la taille de l'ensemble de prédictions en conséquence.

Il a été montré empiriquement que la précision de positionnement était d'environ 2,4 m / probabilité de 75

Ces résultats ont surpassé les algorithmes de Machine Learning sans indince de confiance mis à l'essai sur les mêmes bancs d'essai jusqu'à 20

Les approches présentée dans cet article ne nécessite pas de carte du bâtiment. Cela pourrait être utile de les avoir afin d'avoir des informations supplémentaires pour supprimer les mauvaises prédicitons comme par exemple une personne qui marche dans un mur.

## 1.6 Choix de l'algorithme

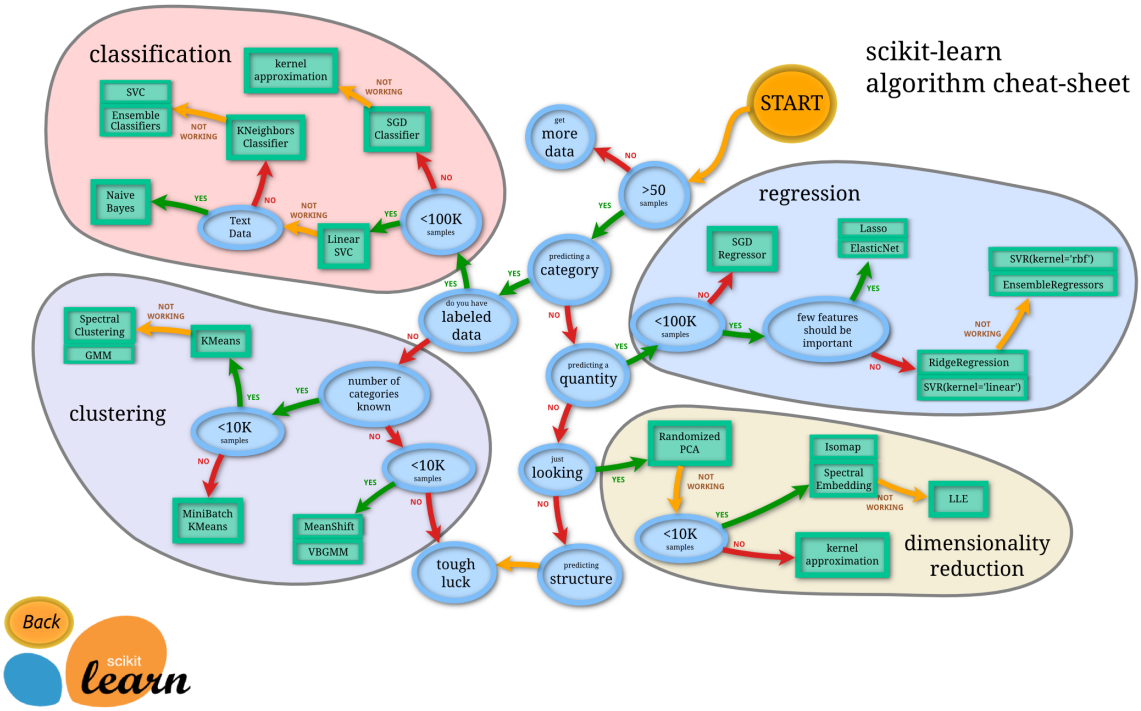


**Figure 1.13 –** Comment choisir un algorithme ML

# 2 Rapport intérmédiaire : 15.10.2018 au 26.10.2018

## 2.1 Objectifs et tâches à réaliser

# 3 Introduction

## 3.1 Contexte

La précision du positionnement intérieur reste quelque chose de très important car il peut être utile dans plusieurs domaines comme : La gestion de stocks, la localisation de personnes âgées dans des homes, la localisation chez eux des personnes possédant un bracelet de "prisonnier", etc...

Les systèmes de localisation basés sur GPS souffrent de la détérioration de la précision et sont presque indisponibles dans les environnements intérieurs. Pour les environnements intérieurs, de nombreuses technologies de systèmes de positionnement ont été conçues sur la base de la vision, de la détection infrarouge ou ultrasonore, des champs magnétiques de la terre, des accéléromètres / gyromètres, des balises BLE ou de la communication WiFi. Bien que la création de ces nouvelles applications ait été couronnée de succès, le coût de ces récepteurs, leur consommation d'énergie et leur limitation aux environnements extérieurs excluent de nombreuses applications.

La figure 3.1 montre un graphique comparatif de la précision de positionnement concernant différentes technologie [LLI].

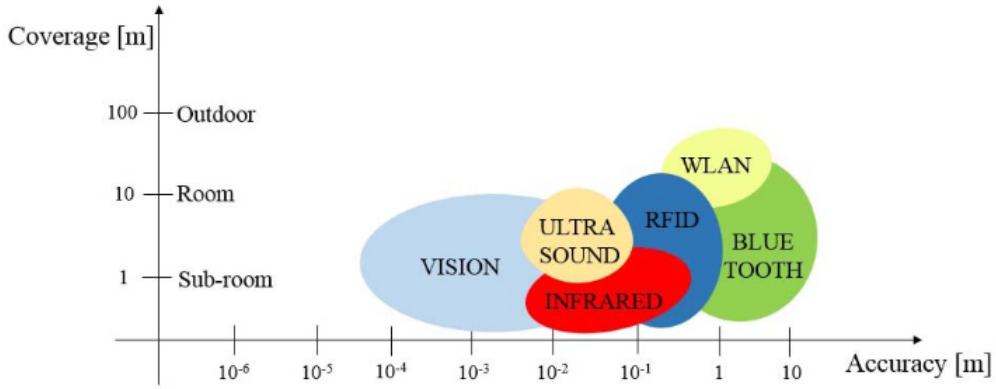


**Figure 3.1 –** Etat de l'art des différentes méthode de positionnement intérieur [LLI]

La géolocalisation avec LoRa est une possibilité séduisante et probablement l'un des meilleurs candidats pour le positionnement intérieur. Le faible coût des infrastructures et des noeuds finaux ainsi que la disponibilité à l'échelle de la ville ou du pays pourraient permettre de nombreuses nouvelles applications. Il n'est donc pas surprenant que les chercheurs et les entités commerciales se soient mis au travail sur ce problème au cours des derniers mois. Cependant, plusieurs défis restent à relever pour qu'un tel système devienne pratique. Premièrement, la précision de localisation en extérieur qui peut actuellement être atteinte avec LoRa est comprise entre 30 et 50 mètres, ce qui n'est pas suffisant pour de nombreuses applications en milieu urbain. Deuxièmement, très peu d'expérience est disponible pour la conception de systèmes de positionnement intérieurs avec LoRa.

La portée en milieu rural est d'environ 15km alors qu'il est de 5km dans un milieu urbain cela grâce à la bonne sensibilité du récepteur (-130dBm). Une chose intéressant est la bande passante qui est plus large que d'autres technologies qui permet de distinguer différents chemins du même signal. Sagemcom ont obtenus des bons résultats au niveau de la précision qui est de environ 4 mètres.[FP17]

L'objectif général de ce projet est de développer les technologies permettant d'améliorer la précision de la géolocalisation en intérieur sur la base de la technologie LoRa.

Le positionnement intérieur reste quelque chose de très intéressant et important car il peut être utile dans plusieurs domaines comme : La gestion de stocks, la localisation de personnes âgées dans des homes, la localisation chez eux des personnes possédant un bracelet de "prisonnier", etc...

## 3.2 Aspect Novateur

Ce travail d'approfondissement doit évaluer une nouvelle approche pour améliorer le positionnement intérieur. En s'appuyant sur les capacités étendues des nouveaux circuits intégrés LoRa, ce projet développera et déploiera un système de localisation capable d'améliorer la précision de la position atteinte par les systèmes de localisation basés sur LoRa existants reposant sur des mécanismes TDOA ou de "ranging". À cette fin, une exploration et une comparaison des différentes techniques "machine learning/deep learning" pour le positionnement basé sur le "fingerprinting" seront effectuées.

L'aspect novateur du projet et d'intégrer un mécanisme d'apprentissage de la position afin d'améliorer la précision.

## 3.3 Objectifs et tâches à réaliser

1. Etudier le cahier des charges
2. Etudier l'état de l'art des techniques à utiliser dans le cadre du projet, en particulier les systèmes de localisation indoor basés sur des techniques d'apprentissage, et réunir une documentation (env. 20
3. Etablir un planning pour l'ensemble du projet.
4. Définir un plan des tests à effectuer.
5. Définir les procédures de test
6. Définir le setup pour la collecte de données de localisation
7. Prise en main de l'environnement de développement pour les phases de training et du test de la technique d'apprentissage retenue (e.g., PyTorch).
8. Implémentation de la solution ML retenue.
9. Tester le système selon le protocole préétabli.
10. Faire des propositions pour améliorer les performances de l'algorithme et, si possible, les implémenter.
11. Rédiger le rapport et documenter l'ensemble du projet.

## 3.4 Structure du rapport

Compléter cette partie en fin de rapport quand la structure est définie

# 4 Etat de l'art

Ce chapitre traite de l'état de l'art sur ce qui existe déjà en matière de "Machine Learning" concernant le positionnement intérieur. Il est important de voir ce qui se fait afin d'économiser un temps précieux pour ne pas partir dans une mauvaise direction et ainsi pouvoir être plus efficace dans la recherche.

Pour ce faire, quatre ouvrages ont été étudiés :

1. A Comparative Study on Machine Learning algorithms for Indoor Positioning [Sin+15]
2. A Machine Learning Approach to Ranging Error Mitigation for UWB Localization [Hen+12]
3. A performance guaranteed indoor positioning system using conformal prediction and the WiFi signal strength [KN17]
4. GPS-free Geolocation using LoRa in Low-Power WANs [FP17]

Les trois premiers traitent du positionnement intérieur aidé avec des algorithmes de "Machine Learning". Le quatrième a été sélectionné car il traite de la gestion de "outliers", c'est à dire comment détecter des points abstraits et qui pourraient fausser les mesures.

## 4.1 Extrait de choses importantes

Cette section permet de mettre en évidence les aspects importants qui ressortent des différentes lectures.

### 4.1.1 Type d'apprentissage

mettre le schéma de scikit learn et expliquer les différentes alternative

### 4.1.2 Algorithme

Dans cette section, quelques algorithmes de classification utilisés dans ces études [FP17] [Hen+12] [KN17] sont brièvement décrits.

#### Decision Tree

L'arbre de décisions et une méthode très connue en "machine learning". Il possède des noeuds de décisions (non-terminal), des branches, et des noeuds feuilles (terminal) qui représentent les caractéristiques, condition et les classes. A chaque noeud de décision on sait quelle branches suivre et lorsque l'algorithme atteint un noeud final, le label contenu dans ce même noeud est retourné comme étant la classe. L'ID3 de Quinlan et son successeur, C4.5, sont les plus populaires parmi les algorithmes d'arbre de décision [JR14].

#### Naïve Bayes

Le classificateur Naïve Bayes [GHP95] basé sur le théorème de Bayes est un algorithme d'apprentissage supervisé [CS13]. Il est robuste aux données bruyantes, facile à construire, affiche une grande précision et rapidité lorsqu'il est appliqué à de grandes bases de données et exécute des modèles de classification plus complexes. Par conséquent, il est largement utilisé dans les tâches de classification. Il calcule la probabilité de chaque attribut dans les données en supposant qu'elles sont également importantes et indépendantes les unes des autres. Cette hypothèse est appelée indépendance conditionnelle de classe [WA13] [SS14].

### Bayesian Network

L'algorithme de réseau bayésien est largement utilisé pour la classification et est basé sur le théorème de Bayes où la probabilité conditionnelle sur chaque nœud est calculée et forme un réseau bayésien. Il s'appelle également réseau de croyance ou réseau occasionnel. Réseau bayésien a deux parties nommées qualitatives et quantitatives, qui sont la structure topologique du réseau bayésien et le tableau de probabilité conditionnelle (CPT), respectivement [DL].

Le réseau bayésien est un graphe acyclique dirigé où chaque nœud représente un attribut des données et un ensemble de distributions de probabilité. Ces distributions donnent les probabilités pour la valeur de chaque nœud étant donné que les parents de nœud.

### K-Nearest Neighbor

Le classificateur K-Nearest Neighbor (K-NN) [DWDMK] est également connu sous le nom de classificateur basé sur la distance qui classe les instances en fonction de leur similarité. C'est l'un des algorithmes les plus populaires de l'apprentissage automatique. C'est un type d'apprentissage paresseux dans lequel la fonction n'est approchée que localement et tout calcul est retardé jusqu'à la classification. Le tuple inconnu dans K-NN est assigné à la classe la plus commune parmi ses K-plus proches voisins. Lorsque K = 1, le tuple inconnu se voit attribuer la classe du tuple d'apprentissage le plus proche dans l'espace des motifs [CAG].

### SMO

L'algorithme d'optimisation séquentielle minimale (SMO - Sequential minimal optimization) [29] est représenté par John C. Platt pour la formation du classificateur de vecteurs de support à l'aide des noyaux polynomiaux ou RBF. C'est l'un des algorithmes les plus courants pour la classification des grandes marges par SVM. Il remplace globalement toutes les valeurs manquantes et transforme les attributs nominaux en attributs binaires. La SVM est une technique de classification basée sur la technologie des réseaux neuronaux utilisant la théorie de l'apprentissage statistique [30]. Il recherche un hyperplan optimal linéaire afin de maximiser la marge de séparation entre la classe positive et la classe négative. En pratique, la plupart des données ne sont pas linéairement séparables ; ainsi, pour rendre la séparation possible, la transformation est effectuée à l'aide d'une fonction du noyau. L'entrée est transformée en un espace caractéristique de dimension supérieure à l'aide d'une cartographie non linéaire [30]. Une décision sur la fonction du Kernel est nécessaire pour implémenter SVM. Le Kernel définit la classe de fonction [31].

[29] J. Platt, "Fast Training of Support Vector Machines using Sequential Minimal Optimization", Advances in Kernel Methods - Support Vector Learning, 1998.

[30] P. Niken, and H. Ohwada, "Applicability of machine-learning techniques in predicting customer defection", IEEE 2014 International Symposium on Technology Management and Emerging Technologies (ISTMET), 2014.

[31] S. M. Obaidullah, K. Roy, and N. Das, "Comparison of different classifiers for script identification from handwritten document", 2013 IEEE International Conference on Signal Processing, Computing and Control (ISPCC), pp.1-6, 2013.

### AdaBoost

AdaBoost (Adaptive Boosting) [32] est un algorithme d'apprentissage d'ensemble. Généralement, il peut être utilisé avec des algorithmes de Machine learning faibles pour améliorer leurs performances. Il est simple à mettre en œuvre, rapide et moins susceptible d'avoir un overfitting. Il améliore les algorithmes de classification instables tels que J48, DecisionStump, etc. L'idée derrière cet algorithme est d'obtenir un classificateur très précis en combinant de nombreux classificateurs faibles. Il fonctionne en exécutant de manière répétée un algorithme d'apprentissage faible donné sur diverses distributions sur les données d'apprentissage, puis en combinant les classificateurs produits par l'apprenant faible en un classificateur

composite unique [33]. Les classificateurs de l'ensemble sont ajoutés un par un, de sorte que chaque classificateur suivant est entrainé sur des données difficiles pour les membres précédents de l'ensemble. Les poids sont définis sur les instances du jeu de données, en suivant une règle selon laquelle les instances difficiles à classer prennent plus de poids. Cette règle conduit les classificateurs ultérieurs à se concentrer sur eux [34].

[32] Y. Freund, and R. E. Schapire, "Experiments with a new boosting algorithm", 3th International Conference on Machine Learning, San Francisco, pp. 148-156, 1996.

[33] R. Shams, and R. E. Mercer, "Classifying Spam Emails Using Text and Readability Features", 2013 IEEE 13th International Conference on Data Mining (ICDM, pp. 657-666, 2013.

[34] S. O. Sharif, L. I. Kuncheva, and S. P. Mansoor, "Classifying encryption algorithms using pattern recognition techniques", 2010 IEEE International Conference on Information Theory and Information Security (ICITIS), , pp. 1168-1172, 2010.

#### Bagging

le Bagging [35] crée des sacs de données de la même taille que le jeu de données d'origine en appliquant une sélection aléatoire à différents sous-ensembles des données d'apprentissage avec de nombreux exemples qui apparaissent plusieurs fois. Ce processus est appelé réplication bootstrap des données d'entrainement. L'idée derrière cette technique est de construire différents classificateurs en utilisant ces sous-ensembles. Chaque sous-ensemble est utilisé pour entrainer un classificateur individuel. Cette approche d'ensemble utilise le nombre de classificateurs a priori [35].

[35] L. Breiman, "Bagging predictors", Machine Learning. vol. 24, no. 2, pp.123-140, 1996.

#### support vector machine(SVM) regression

#### Gaussian process (GP)

### 4.1.3 Outliers [FP17]

Cette section donne un apperçu de la manière de traiter les "outliers - valeurs aberrantes", c'est-à-dire les points qui ne sont pas cohérent lors d'une mesure. Selon Barnet et Lewis [VT94], un "outliers" est définit comme étant une observation qui semble incompatible avec le reste d'un ensemble de données. Garder un "outliers" dans une set de données peut amener à de mauvais résultats, il est donc important de les détecter correctement. Il existe différentes méthodes pour déterminer ces "outliers" :

1. Grubbs' test : Détecte un "outliers" en supposant une distribution normale.
2. Tietjen-Moore test : C'est une généralisation de Grubbs' test pour détecter de multiple outliers. Il a cependant un inconvénient, il est nécessaire de connaitre le nombre exact d'ouliers.
3. Generalized Extreme Studentized Deviate (ESD) : C'est également une généralisation du test Grubbs' mais il n'est pas nécessaire de connaitre à l'avance le nombre d'ouliers. Ce test nécessite uniquement une limite supérieure pour le nombre suspect d'outliers.[Esd]

## 4.2 Comparaison

## 4.3 Choix

# Bibliographie


[LLI] Mainetti Luca, Patrono Luigi et Sergi Ilaria. *A Survey on Indoor Positioning Systems*. Rapp. tech. University of Salento Lecce, ITALY (cf. p. 17).

[FP17] Bernat Carbones Fargas et Martin Nordal Petersen. « GPS-free Geolocation using LoRa in Low-Power WANs ». In : *Proceedings of 2017 Global Internet of Things Summit (GIoTS)* (2017) (cf. p. 17, 19, 21).

[Sin+15] Bozkurt Sinem et al. *A Comparative Study on Machine Learning algorithms for Indoor Positioning*. Rapp. tech. This work is supported by The Scientific et Technological Research Council of Turkey (TUBITAK) under grant number 1130024, 2015 (cf. p. 19).

[Hen+12] Wymeersch Henk et al. *A Machine Learning Approach to Ranging Error Mitigation for UWB Localization*. Rapp. tech. Belgian American Education Foundation, the Charles Stark Draper Laboratory Robust Distributed Sensor Networks Program, the Office of Naval Research Young Investigator Award N00014-03-1-0489, the National Science Foundation under Grants ANI-0335256 and ECCS-0636519., 2012 (cf. p. 19).

[KN17] Khuong et Nguyen. « A performance guaranteed indoor positioning system using conformal prediction and the WiFi signal strength ». In : *Journal of Information and Telecommunication, 1 :1, 41-65, DOI : 10.1080/24751839.2017.1295659* (2017). URL : https://doi.org/10.1080/24751839.2017.1295659 (cf. p. 19).

[JR14] Quinlan J. R. « C4. 5 : programs for machine learning ». In : *Elsevier* (2014) (cf. p. 19).

[GHP95] John G. H. et Langley P. « Estimating Continuous Distributions in Bayesian Classifiers ». In : *11th Conference on Uncertainty in Artificial Intelligence, pp., 338-345* (1995) (cf. p. 19).

[CS13] Anuradha (23) C. et Dhall S. « Software Defect Prediction Using Supervised Learning Algorithm and Unsupervised Learning Algorithm ». In : (2013) (cf. p. 19).

[WA13] Yotsawat W. et Srivihok A. « Inbound tourists segmentation with combined algorithms using K-Means and Decision Tree ». In : *10th International Joint Conference on Computer Science and Software Engineering (JCSSE), pp.189-194* (2013) (cf. p. 19).

[SS14] Ureerat S. et P. Singsri. « The classifier model for prediction quail gender after birth based on external factors of quail egg ». In : *IEEE 11th International Joint Conference on Computer Science and Software Engineering (JCSSE)* (2014) (cf. p. 19).

[DL] Yang D. et Jin-lin L. « Research on personal credit evaluation model based on bayesian network and association rules ». In : *International Conference on Wireless Communications, Networking and Mobile Computing* () (cf. p. 20).

[DWDMK] Aha D. W., Kibler D. et Albert M. K. « Instance-based learning algorithms ». In : *Machine Learning, vol. 6, pp., 37-66* () (cf. p. 20).

[CAG] Shah C. et Jivani A. G. « Comparison of data mining classification algorithms for breast cancer prediction ». In : *Fourth International Conference on Computing, Communications and Networking Technologies (ICCCNT)* () (cf. p. 20).

[VT94] Barnett V. et Lewis T. « Outliers in Statistical Data ». In : *3rd ed. Wiley Series in Probability and Mathematical Statistics* (1994). URL : https://doi.org/10.1080/24751839.2017.1295659 (cf. p. 21).

[Esd] « lien concernant les ESD ». In : (). URL : https://www.itl.nist.gov/div898/handbook/eda/section3/eda35h3.htm (cf. p. 21).