Autonomous University of Baja California
School of Engineering

6th International Electronic Conference on Sensors and Applications

Indoor Localization through Mobile Participatory Sensing and Magnetic Field

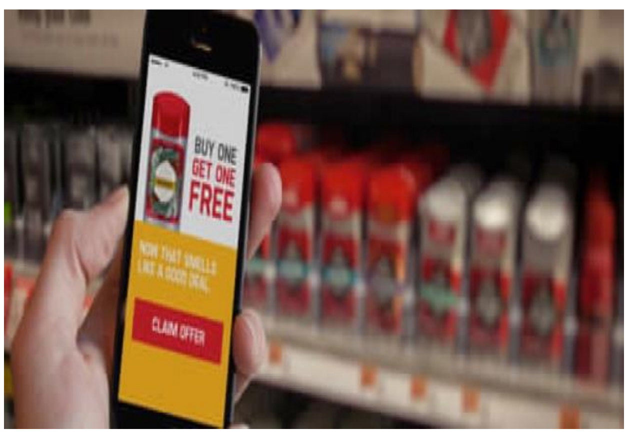
Presented by:
Juan Pablo García Vázquez
pablo.garcia@uabc.edu.mx

Content

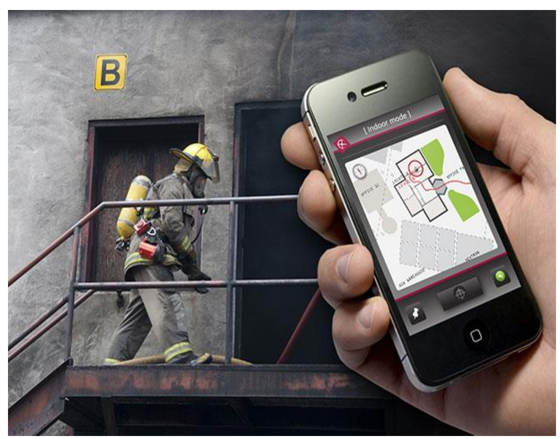
- Problem Context
- Related Work
- Research Aims
- Research Methodology
- Results
- Conclusions
- Future Work

Introducction

- The **location** is the place where an object or person is located [1].
- It can be expressed in form: **Physical** (G: M: S) or **Symbolic** (e.g. living room, dining room, etc.) [2].
- **Importance:** Applications that offer location-based services, for example:



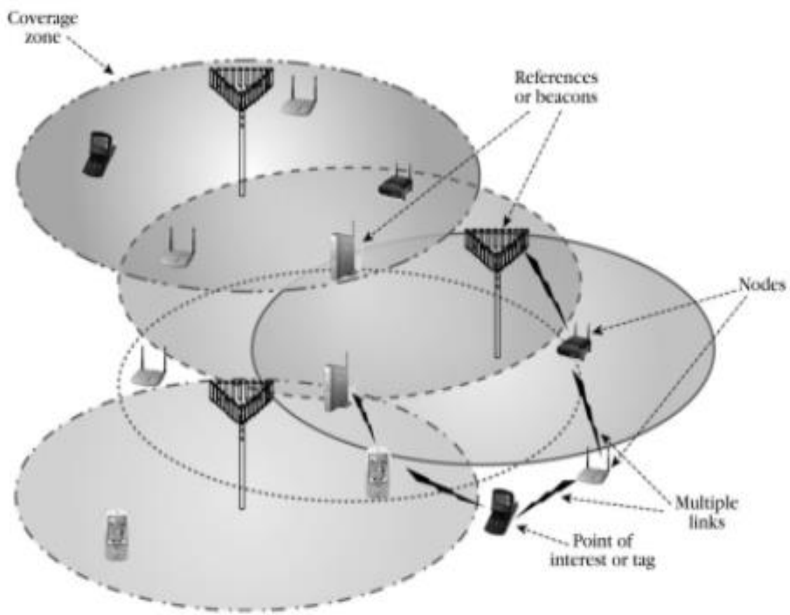
ByteLight [3]



Firefighter [4]

Telocate Assist[5]

Technological Approaches for Localization (1/3)



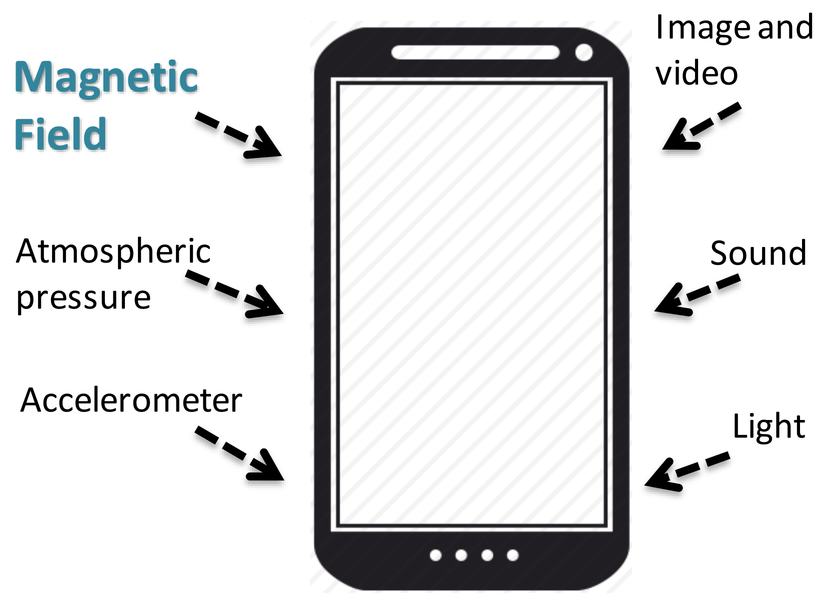
In indoor environments GPS systems are limited - errors up to 50 mts-

[6]http://www.theplace4change.com/blog/2014/09/11/que-son-los-beacons-y-cual-es-su-potencial

Indoor Localization Approaches (2/3)

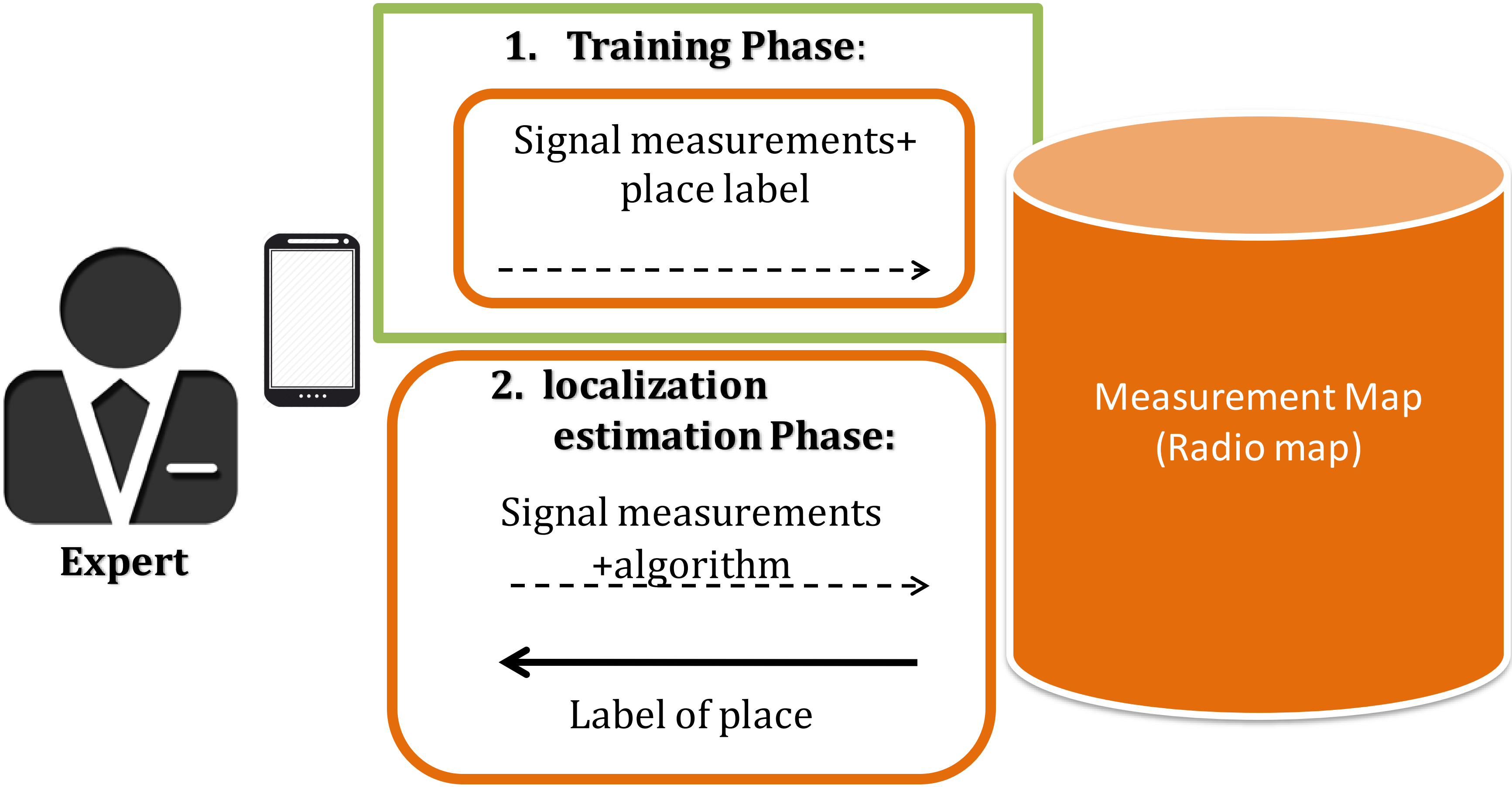
Bluetooth

They require transmitting and receiving devices in the environment



They use the environmental signals available in the environment.

Fingerprint



[8] Taheri, A., Singh, A., Emmanuel, A. Location fingerprinting on infrastructure 802.11 wireless local area networks (WLANs) using Locus. In 29th Annual IEEE International Conference on Local Computer Networks, pp. 676-683, 2004

Fingerprint Requirements

- The measurement is performed by an "expert".
  - Has knowledge of the signal of interest (measuring instruments, propagation, units, etc.)
- If the environment changes, the radio map must be updated.
  - Limit its implementation.
- The data collection stage involves high cost and time.
  - For the experience of the expert and specialized team.

Participatory Sensing

- Paradigm that empowers multiple users to contribute to data detected or generated on their mobile devices[9].

Advantages

- *No expert is required for data collection*
- *Less time and cost for measure map*
- *Dynamic update of measure map*

[9] Estellés-Arolas, E., González, F. Towards an integrated crowdsourcing definition. *J. of Information science*, 38(2):189-200, 2012.

Indoor Localization and Participatory Sensing

Systems for the automatic construction indoor Layouts

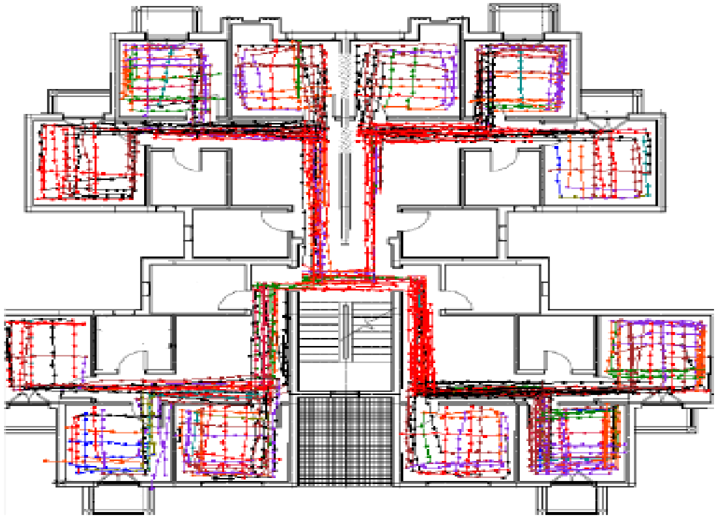


Figure 2. CrowdInside
(Fuente: Alzantot et al. [11])

The projects do not provide evidence of the use of participatory sensing by collecting the magnetic field signal to generate a radio map used to estimate the location of an individual indoors.

Research Questions

- What is the accuracy that can be achieved indoors using magnetic field and participatory sensing?
- Which of the classification algorithms used with magnetic field offers the best accuracy?

Research Aim

Determine the feasibility of applying the participatory sensing as a stage of data collection in the technique of the signal fingerprint for the development of an indoor location system.

Specific Research Aims

- Generate a radio map with magnetic field strength measurements collected through Participatory Sensing

- Create a predictive model to estimate the location of an individual in an indoor environment using machine learning algorithms.

- Evaluate the efficiency of the proposed predictive model to estimate the location of an individual in an indoor environment.

Research Methodology

Data Collection → Data Processing → Indoor Model Generation → Indoor Model Testing

- Based on the reference model for the development of data mining projects.
- *CRISP-DM (Cross Industry Standard Process for Data Mining)* [13].

[13] Wirth, R. CRISP-DM : Towards a Standard Process Model for Data Mining. Proc. of the 4th Intl. Conf. on the Practical Application of Knowledge Discovery and Data Mining, 24959, 29–39, 2000.

Research Methodology

Data Collection → Data processing → Indoor Model Generation → Indoor Model Testing

Objective: To generate the experimental data set of the magnetic field strength necessary to create the indoor location radio map.

- A software component for Android mobile devices was developed that used the magnetometer sensor.

Research Methodology

Data Collection → **Data Processing** → Indoor Model Generation → Indoor Model Testing

Objective: Provide structure and format to data from distributed open collaboration.

- Through the use of the data analysis tool: Opal Convert and Data Normalization.

Research Methodology

Data Collection → Data Processing → **Indoor Model Generation** → Indoor Model Testings

Objective: Identify the relationships between the different characteristics of the magnetic field intensity associated with a location through a predictive model.

- Models were generated using classification algorithms, such as: Decision trees (J48), Nearest neighbors (k-NN), Naïve Bayes.

Research Methodology

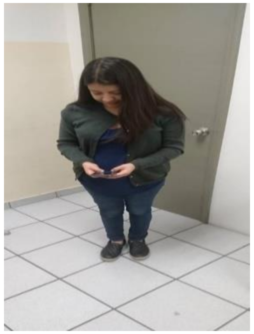
Data Collection → Data Processing → Indoor Model Generation → **Indoor Model Testing**

Objective: To evaluate the predictive models generated with data from distributed open collaboration.

- Metrics were used to measure the accuracy of the algorithm (e.g. confusion matrices)

Phase of data collection

Participant Selection → Training → Data Collection



Device	Magnetometer
Sony C6906	AK8963
Sony d6503	Ak09911
Google Nexus 4	LGE
Samsung SM-G531H	YAS537
Google Nexus 7	Invense MPL
Verizon SM_G900V	AK09911C
Sony E5803	Invense Inc.

Data Collection

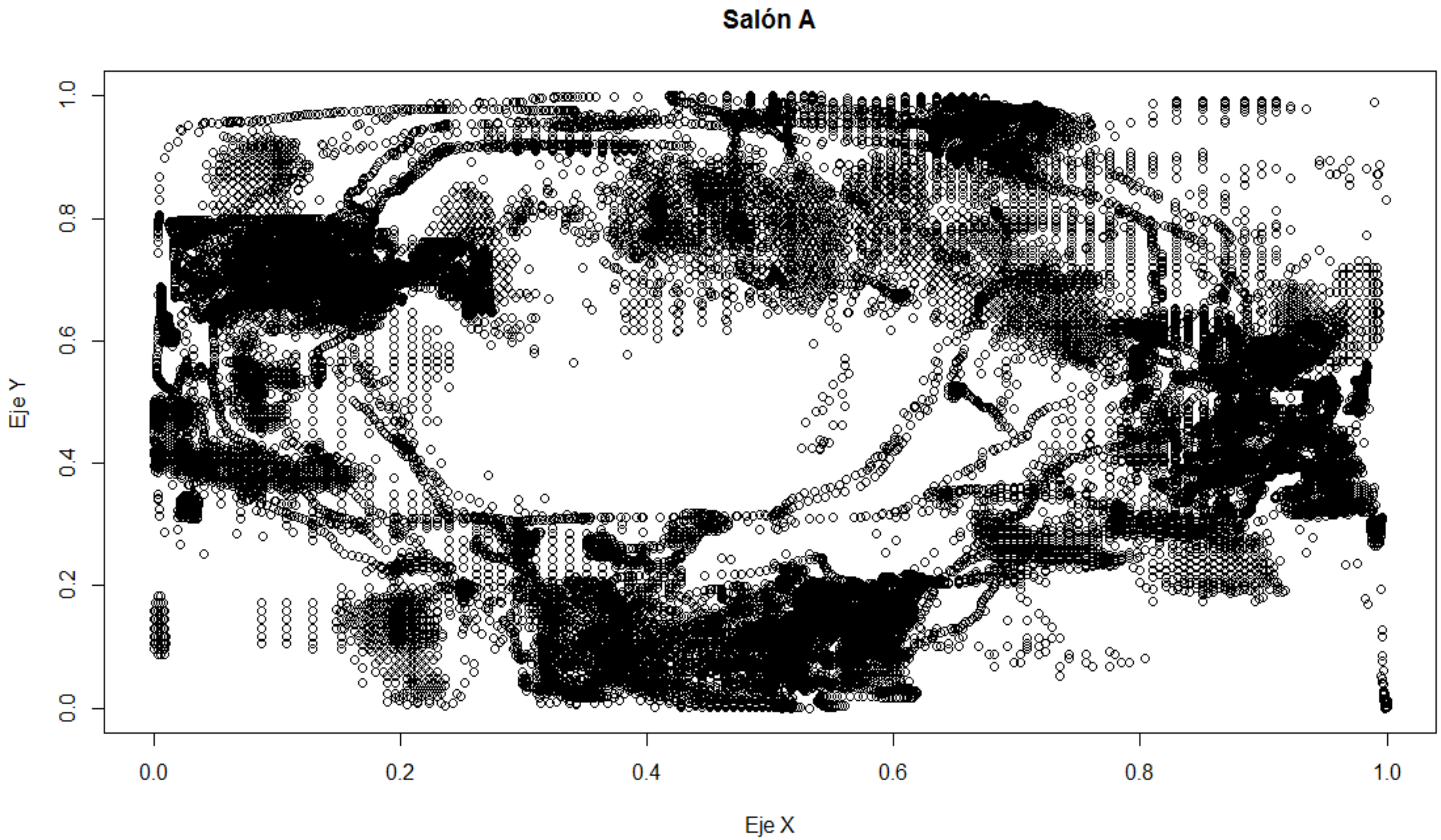
A=64 m^2
B=32m^2
C, D y E = 55.2 m^2

Fig 4. Layout of Computer System Laboratory

Table . Device Model used in data collection phase

Mobile Device	Magnetometer Model
Sony C6906	AK8963
Sony d6503	Ak09911
Google Nexus 4	LGE
Samsung SM-G531H	YAS537
Google Nexus 7	Invense MPL
Verizon SM_G900V	AK09911C
Sony E5803	Invense Inc.

Data Collection

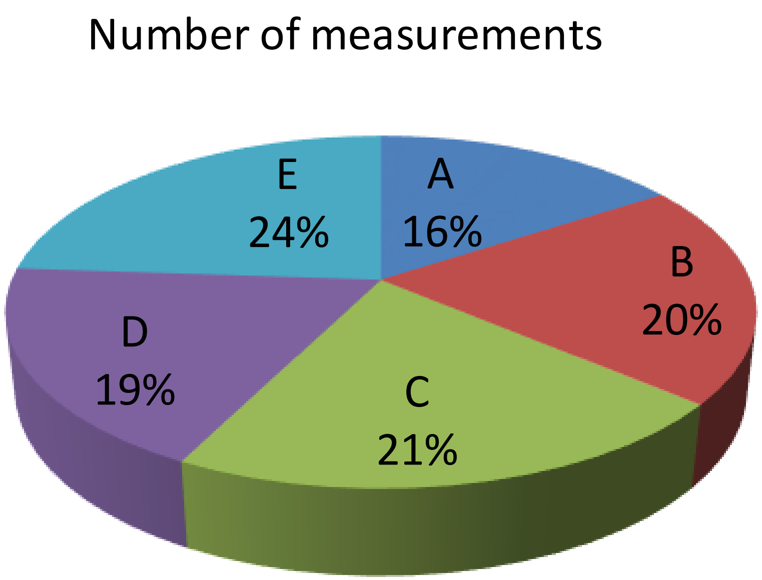


Results: Data Collection

- A total of 701,714 magnetic field intensity measurements were obtained using various devices (see Table 1).

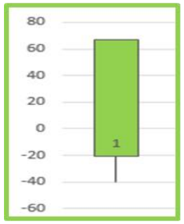
Table 1. Magnetic field intensity measurements per room

Room	Number of measurements
A	110,679
B	143,352
C	149,135
D	130,708
E	167,840

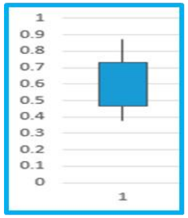


Phase: Data Processing

Normalization



Norm	X	Y	Z
71.45603	-26.7822	-57.5988	32.727
71.6735	-27.275	-58.1054	31.8893
71.9737	-26.7822	-58.612	32.0556
72.70242	-26.6189	-59.1186	32.8948
71.67288	-27.4383	-57.936	32.0556



norm	x	y	z
0.46889872	0.87291086	0.37601084	0.73421909
0.46506584	0.87291086	0.40091217	0.75491366
0.46763595	0.87291086	0.37601084	0.75491366
0.45794481	0.84675914	0.36376692	0.78570049
0.46512774	0.84675914	0.34919059	0.7691582
0.46820295	0.85529893	0.34919059	0.7691582

Data Partition

- The data was randomly divided into two subsets:

Training	70%
Test	30%

[17] C. E. Galván-Tejada, J.P. García-Vázquez, J.I. Galván-Tejada, J. R. Delgado-Contreras, R. Brena, “Infrastructure-less indoor localization using the microphone, magnetometer and light sensor of a smartphone”. Sensors, 15(8), 20355-20372, July 2015.

k-NN: Confusion Matrix

Confusion matrix for the generated model with a value of k = 3

Referencia

Predicction	LAB A	LAB B	LAB C	LAB D	LAB E	Total	%CORRECT	%INCORRECT
LAB A	**37,688**	288	608	349	366	39,299	95.90%	4.09%
LAB B	281	**32,512**	166	177	84	33,220	97.86%	2.13%
LAB C	552	169	**48,961**	359	219	50,260	97.41%	2.58%
LAB D	353	149	417	**43,402**	447	44,768	96.94%	3.05%
LAB E	338	85	200	453	**41,889**	42,965	97.49%	2.30%
					TOTAL	**210,512**	97.12%	2.83%

The confusion matrix shows that of the 39,299 data of the LAB A class, 37,688 data were correctly classified with 95.90% and 4.09% incorrectly.

k-NN: Confusion Matrix

Confusion matrix for the generated model with a value of k = 6

Reference

Prediction	LAB A	LAB B	LAB C	LAB D	LAB E	TOTAL	%CORRECTAS	%INCORRECTAS
LAB A	**37,345**	210	736	288	319	38,898	96.01%	4.09%
LAB B	306	**32,886**	193	228	115	33,728	97.50%	2.50%
LAB C	529	122	**48,599**	311	193	49,754	97.70%	2.33%
LAB D	427	103	503	**43,471**	427	44,931	96.80%	3.20%
LAB E	370	75	258	535	**41,965**	43,203	97.13%	2.69%
					TOTAL	**210,512**	97.03	2.96%

In the confusion matrix it is observed that of the 38,898 data of the LAB A class, 37,345 data were correctly classified in 96.01% and 4.09% incorrectly

k-NN: Confusion Matrix

Confusion matrix for the model generated with a value of k = 9

Prediction	Reference LAB A	LAB B	LAB C	LAB D	LAB E	TOTAL	%CORRECT	%INCORRECT
LAB A	**37,507**	284	693	415	389	**39,288**	**95.46%**	**4.53%**
LAB B	266	**32,531**	147	170	95	33,209	97.95%	2.04%
LAB C	725	180	**48,791**	397	265	50,358	96.88%	3.11%
LAB D	386	127	476	**43,287**	485	44,761	97.70%	3.29%
LAB E	328	81	245	471	**41,771**	42,896	97.37%	2.43%
					TOTAL	**210,512**	**96.88%**	**3.11%**

In the confusion matrix it is observed that of the 39,288 data of the LAB A class, 37,507 data were correctly classified in 95.46% and 4.53% incorrectly

J48: Confusion Matrix

A decision tree was built from the 701,714 instances collected (70% for training and 30% for evaluation).

Reference

Prediction	LAB A	LAB B	LAB C	LAB D	LAB E	TOTAL	%CORRECT	%INCORRECT
LAB A	**37,498**	300	666	406	342	39,212	95.63%	4.37%
LAB B	258	**32,564**	150	156	75	33,203	98.08%	1.92%
LAB C	709	169	**48,779**	435	260	50,352	96.88%	3.12%
LAB D	332	175	435	**43,319**	479	44,740	96.82%	3.27%
LAB E	378	91	263	497	**41,776**	43,005	97.14%	2.86%
					TOTAL	**210,512**	96.91%	3.09%

In the confusion matrix it is observed that of the 39,212 data of the LAB A class, 37,498 data were correctly classified in 95.63% and 4.37% incorrectly

Naïve Bayes: Confusion Matrix

The accuracy of the model obtained for this Naïve Bayes algorithm was: 50.77% accuracy with a Kappa of 0.3838.

Referencia

Prediction	LAB A	LAB B	LAB C	LAB D	LAB E	TOTAL	%CORRECT	%INCORRECT
LAB A	**12,688**	7,278	4,092	7,563	7,591	39,212	32.35%	67.64%
LAB B	4,602	**1,778**	3,172	3,199	4,452	33,203	53.54%	46.45%
LAB C	5,418	4,799	**20,179**	14,126	5,830	50,352	40.07%	59.95%
LAB D	5,192	3,036	5,030	**28,265**	3,217	44,740	63.17%	36.82%
LAB E	6,279	2,016	3,913	2,954	**27,843**	43,005	64.74%	30.56%
					TOTAL	**210,512**	50.77%	48.28%

In the confusion matrix, we observe that of the 39,212 data of the LAB A class, 12,688 data were correctly classified in 32.35% and 67.64% incorrectly

Results: Algorithms Performance

Algorithm	Instances correctly classified in %	Instances incorrectly classified in %	Kappa Coefficient in %
KNN (k=6)	97.03 %	2.96 %	96.27 %
KNN (k=3)	97.12 %	2.88 %	96.39 %
KNN (k=9)	96.88 %	3.11 %	91.89 %
J48	96.91 %	3.09 %	96.17 %
Naïve Bayes	50.79 %	49.21 %	38.34 %

The best performing models were obtained from KNN algorithms with k = 3 and J48, with an accuracy of 97.12% and 96.91% respectively to locate an individual.

The model with the lowest performance was obtained with the naive Bayesian algorithm (Naïve Bayes), since it offers an accuracy of 50.79% and a concordance level of 0.3834.

Conclusions

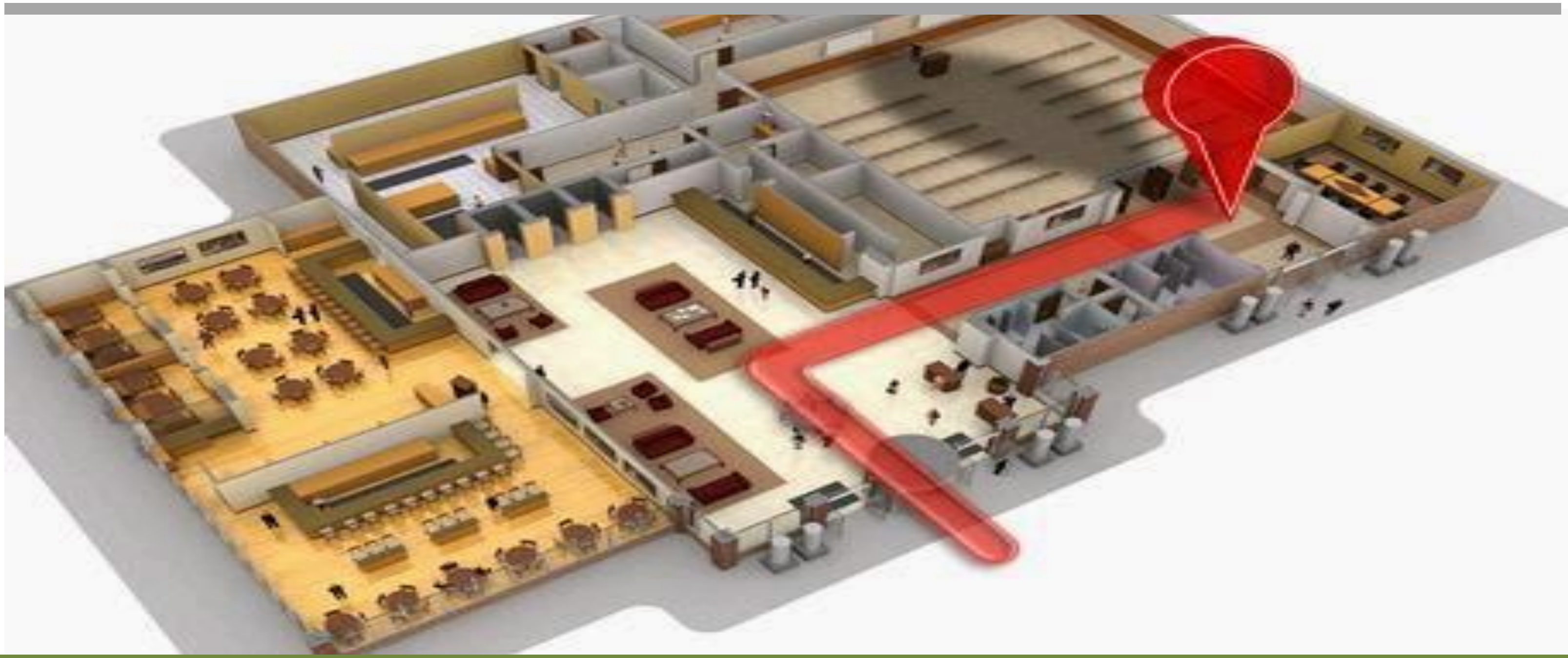
- It is possible to generate a radio map and predictive model with magnetic field strength data collected through the participatory sensing.
- The classification algorithm affects the results of estimating the location of an individual indoors.

Future work

- Explore more robust classification techniques such as SVM, Neural Networks, Random Forest.

- Use other signal features of the magnetic field signal such as time , frequency and statistical.

Autonomous University of Baja California
School of Engineering



Thanks for your attention: Questions or Comments

pablo.garcia@uabc.edu.mx