ITU-ML5G-PS-016: LOCATION ESTIMATION USING RSSI OF WIRELESS LAN

Team Name:
SSN_ITU

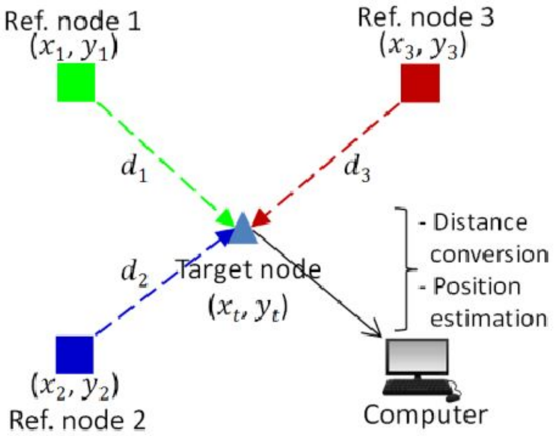
Project Students: Charu Jain
Indu Subramanian
Meghna Govind
N Venkateswaran

OBJECTIVE

1. To propose AI/ML aided localization, utilizing received signal strength indicator (RSSI) observed at the terminal with high accuracy as that of GPS-based localization.
2. Exploit the relationship between distance and RSSI
3. To find latitude and longitude of the nodes of a Wireless network

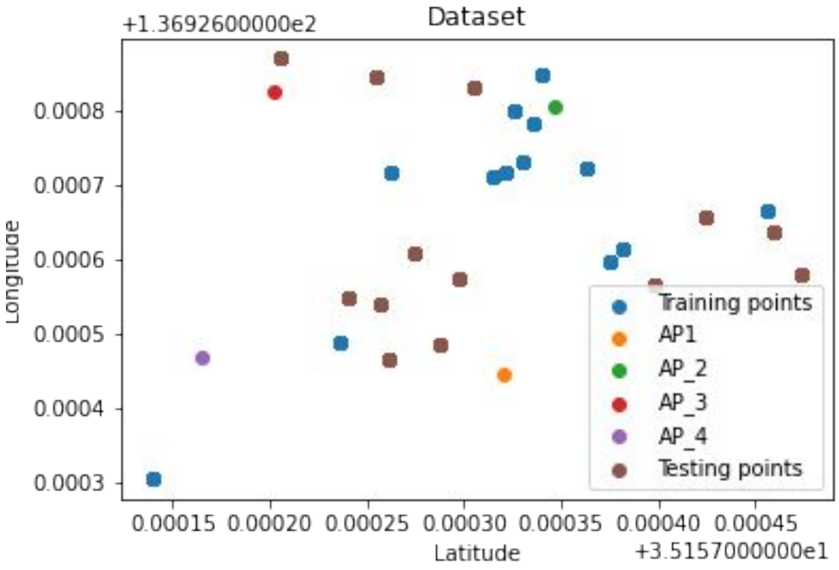
KEY ISSUES WITH RSSI

1. Highly prone to shadowing effects
2. Multipath fading affects the distance-RSSI relationship

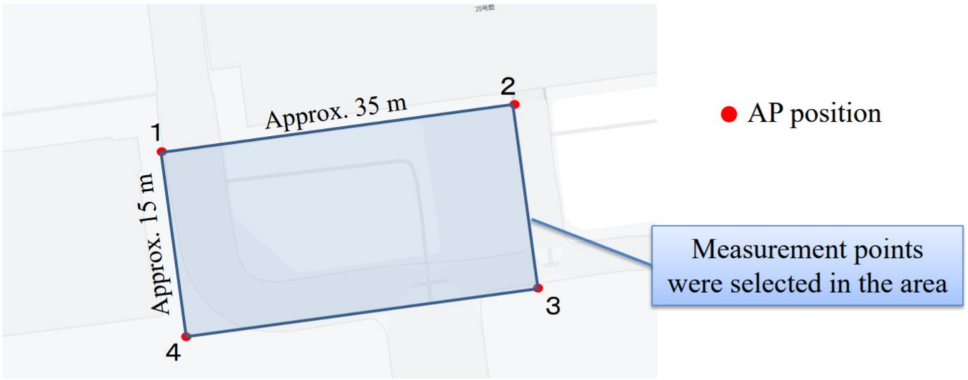


DATASET ANALYSIS

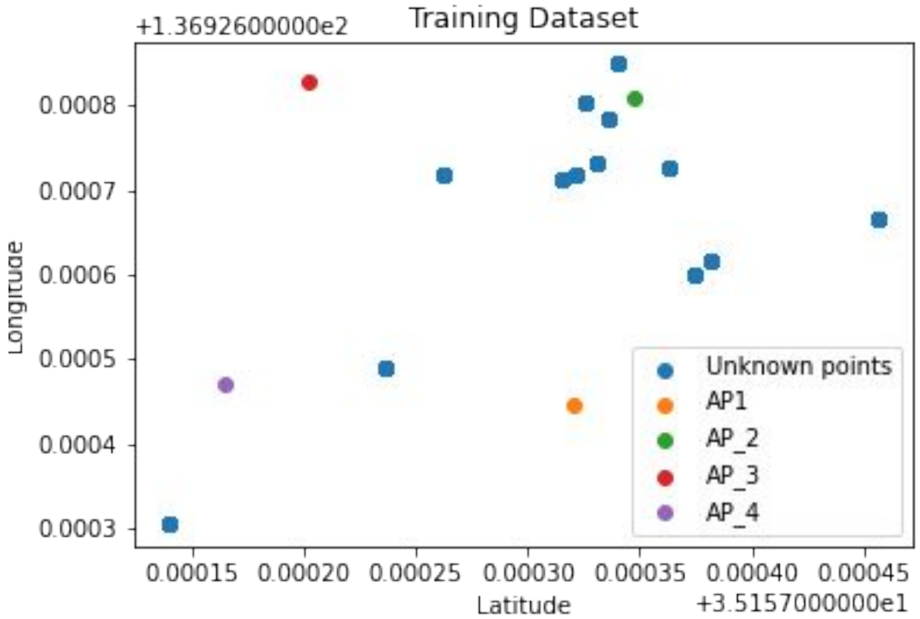
TOPOGRAPHY

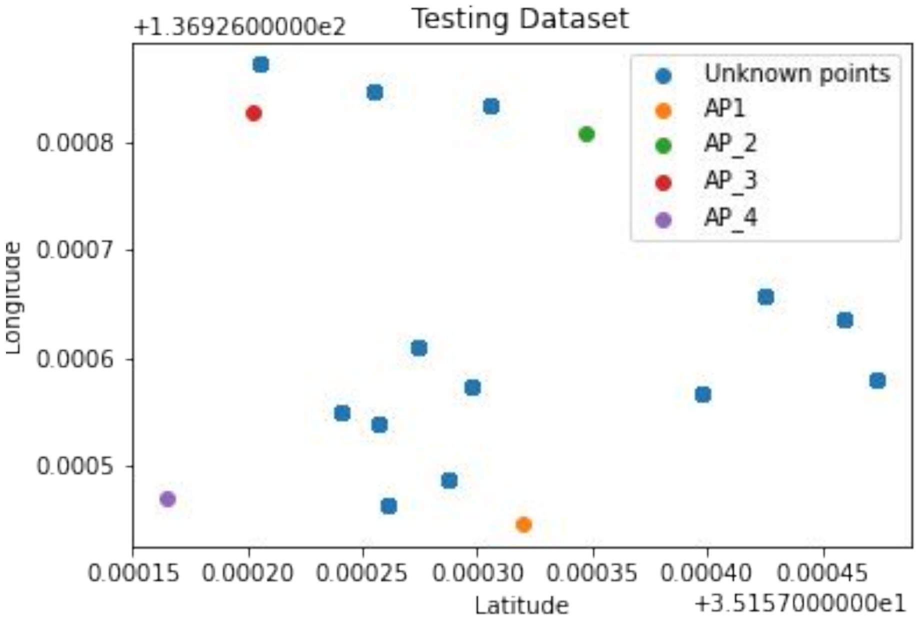


① 4 access points (APs) were located at the location estimation area
② Locations of APs and measurement points were obtained by GPS
③ RSSIs were measured at each measurement point



VISUALISING POSITION OF DATA





INFORMATION GIVEN

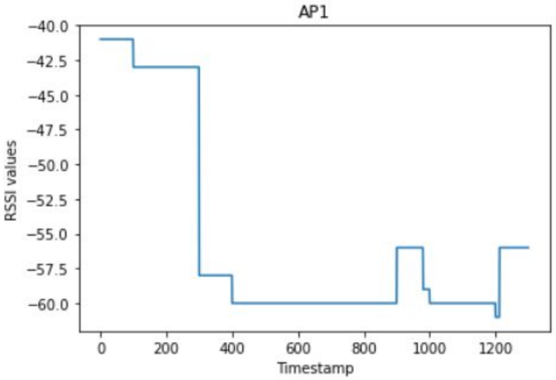
- 4 Access Points
- 13 unknown nodes.
- Latitude, Longitude of APs
- Channel number of each AP
- Timestamps
- RSSI (dBm)

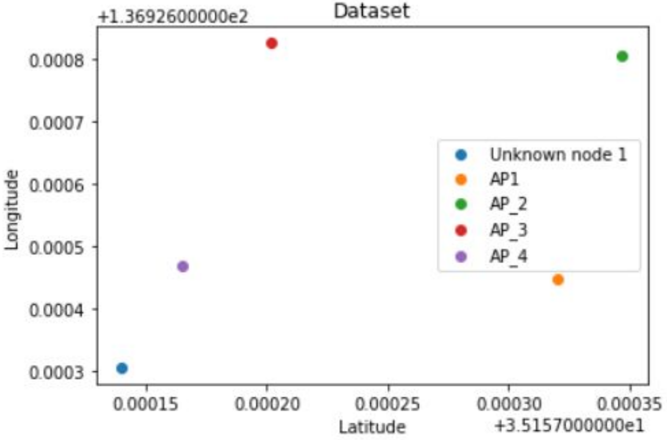
	No.	TimeStamp(UNIX)	Latitude	Longitude	SSID	Channel	RSSI(dBm)
0	1	1631687696	35.157140	136.926306	1	11	-41
1	2	1631687697	35.157140	136.926306	1	11	-41
2	3	1631687697	35.157140	136.926306	1	11	-41
3	4	1631687697	35.157140	136.926306	1	11	-41
4	5	1631687697	35.157140	136.926306	1	11	-41
...	...	...	...	...	...	...	...
5195	5196	1631688763	35.157236	136.926489	4	1	-52
5196	5197	1631688763	35.157236	136.926489	4	1	-52
5197	5198	1631688763	35.157236	136.926489	4	1	-52
5198	5199	1631688763	35.157236	136.926489	4	1	-52
5199	5200	1631688763	35.157236	136.926489	4	1	-52

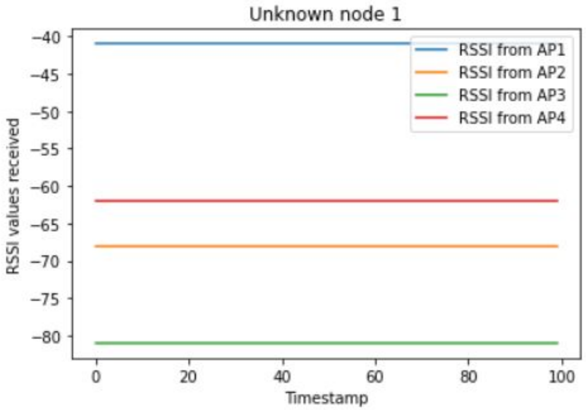
AP SSID	Latitude	Longitude	Hight Difference(m)	Indoor/Outdoor
1	35.1573202	136.926447	0	Outdoor
2	35.157347	136.9268074	0	Outdoor
3	35.1572018	136.9268269	0	Outdoor
4	35.157165	136.9264698	0	Outdoor

DATASET
INFERENCE

- Regression problem
- Sampling period is 100 sec for an unknown node from one AP, i.e an unknown node observes 100 RSSI values from one AP in this period.
- Does not always satisfy the distance-RSSI relationship
- RSSI values to multiple unknown nodes remain the same even though they are in different positions
- Less RSSI values (data points) for a complex environment
- Hence we need to model all these inconsistencies in the dataset







PROPOSED SOLUTION

LET'S CHECK OUT THE SOLUTION

TRAINING
BLOCK DIAGRAM

TimeStamp(UNIX)	Latitude	Longitude	SSID	Channel	RSSI(dBm)	AP_x	AP_y
1631687696	35.157140	136.926306	1	11	-41	35.158465	136.924283
1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283

	AP_1	AP_2	AP_3	AP_4	AP1_x	AP1_y	AP2_x	AP2_y	AP3_x	AP3_y	AP4_x	AP4_y
0	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
2	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
3	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
4	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835

TESTING
BLOCK DIAGRAM

No.	TimeStamp(UNIX)	Latitude	Longitude	SSID	Channel	RSSI(dBm)	AP_x	AP_y
1	1631689137	35.157261	136.926465	1	11	-41	35.158465	136.924283
2	1631689137	35.157261	136.926465	1	11	-41	35.158465	136.924283
3	1631689138	35.157261	136.926465	1	11	-41	35.158465	136.924283
4	1631689138	35.157261	136.926465	1	11	-41	35.158465	136.924283
5	1631689138	35.157261	136.926465	1	11	-41	35.158465	136.924283

AP_1	AP_2	AP_3	AP_4	AP1_x	AP1_y	AP2_x	AP2_y	AP3_x	AP3_y	AP4_x	AP4_y
-41	-70	-73	-64	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
-41	-70	-73	-64	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
-41	-70	-73	-64	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
-41	-70	-73	-64	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
-41	-70	-73	-64	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835

DATA

PREPROCESSING STEPS

01. DATA RESTRUCTURING

No.	TimeStamp(UNIX)	Latitude	Longitude	SSID	Channel	RSSI(dBm)	AP_x	AP_y
1	1631687696	35.157140	136.926306	1	11	-41	35.158465	136.924283
2	1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
3	1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
4	1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
5	1631687697	35.157140	136.926306	1	11	-41	35.158465	136.924283
...	...	...	...	...	...	...	...	...
5196	1631688763	35.157236	136.926489	4	1	-52	35.158931	136.923835
5197	1631688763	35.157236	136.926489	4	1	-52	35.158931	136.923835
5198	1631688763	35.157236	136.926489	4	1	-52	35.158931	136.923835
5199	1631688763	35.157236	136.926489	4	1	-52	35.158931	136.923835
5200	1631688763	35.157236	136.926489	4	1	-52	35.158931	136.923835

	AP_1	AP_2	AP_3	AP_4	AP1_x	AP1_y	AP2_x	AP2_y	AP3_x	AP3_y	AP4_x	AP4_y
0	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
2	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
3	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
4	-41	-68	-81	-62	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
...	...	...	...	...	...	...	...	...	...	...	...	...
1295	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1296	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1297	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1298	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835
1299	-56	-70	-70	-52	35.158465	136.924283	35.158889	136.92435	35.15851	136.923762	35.158931	136.923835

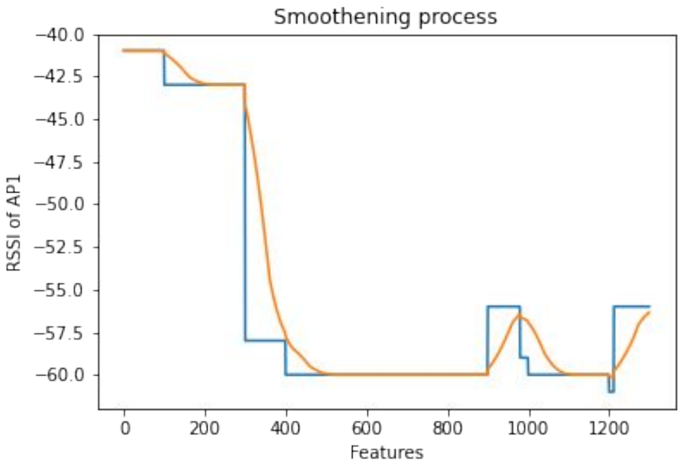
ORIGINAL

5200 rows with 8 features

AFTER RESTRUCTURING

1300 rows with 12 features

02. DATA SMOOTHENING

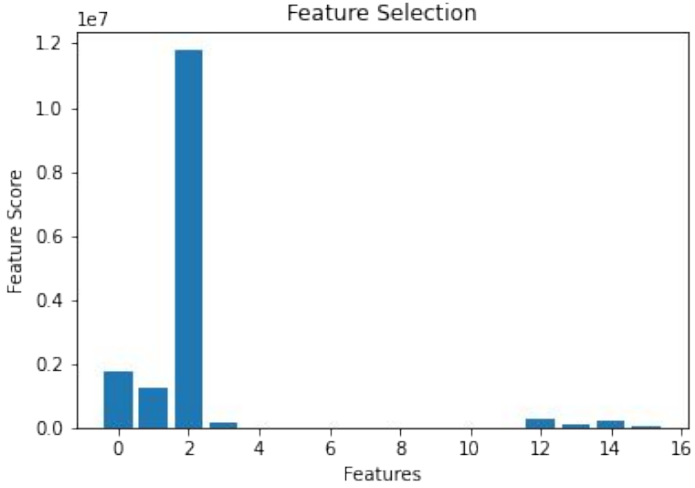


- The RSSI values observed by the target nodes are highly dynamic. And in some cases, the RSSI values are similar for 2 or more target nodes, which hampers accurate learning process of the model.
- This in turn weighs down the accuracy of the model.
- Therefore, we formulated a data smoothening method called the windowing method.
- The process is to select a window size ‘n’ and to determine the average of these values. The weightage of this average value on the current RSSI values is determined by ‘α’.

CURR_VAL = CURR_VAL*α + (1-α)*AVG_VAL

α = 0.2 , to give more weightage to the average of previous values.

03. FEATURE SELECTION



FEATURE 7, 8 AND 9 ARE REMOVED

- Irrelevant or partially relevant features can negatively impact model performance.
- Having irrelevant features in your data can decrease the accuracy of the model.
- Hence we select only those features which contribute most to the output that we are interested in.
- Correlation is a measure of how 2 variables change together.
  SelectKBest() from scikit-learn machine library uses correlation statistics to provide scores for all features
- The scores that are in the negative range or zero are neglected and the the scores that fall under the positive range are considered.

04. DATA AUGMENTATION

- We devised a technique that is used to increase the amount of already existing data.
- We utilized a factor “num”, which determined by what amount the dataset would increase. Based on this factor, rows belonging to a particular target node were chosen to be appended to the same node’s datapoints.
- For every unique target node, “num” number of extra rows are generated.
- Reduces the risk of overfitting and the accuracy on unseen data can be improved.

For the given data, we have used num=10000

- Shape of training dataset : (1300, 18)
- Shape of training dataset after augmenting: (130000, 18)

EXTRA TREES REGRESSION MODEL

EXTRA TREES REGRESSION MODEL

- The Extra Trees algorithm works by creating a large number of unpruned decision trees from the training dataset.
- Predictions are made by averaging the prediction of the decision trees in the case of regression.
- An ensemble method
- Provided by sklearn library

```
def extraTrees_LatModel(X_train, ytrain_lat):

    extra_trees_regressor_lat = ExtraTreesRegressor()
    extra_trees_regressor_lat.fit(X_train,ytrain_lat)

    return extra_trees_regressor_lat
```

```
def extraTrees_LonModel(X_train, ytrain_lon):

    extra_trees_regressor_lon = ExtraTreesRegressor()
    extra_trees_regressor_lon.fit(X_train,ytrain_lon)

    return extra_trees_regressor_lon
```

WHY REGRESSION MODEL

- Regression is a type of supervised learning task. It is used in cases where the value to be predicted is continuous.
- Regression is used to predict the value of the dependent variable for which some information concerning the explanatory variables is available, or in order to estimate the effect of some explanatory variable on the dependent variable.
- As we need to predict the Latitude and Longitude of the unknown node, given a set of dependent features, we have used **Regression**
- We have used **Extra Trees regression model** to model the given ITU Dataset

RESULTS

LET'S SEE HOW THE MODEL WORKS!

ERROR IN MODELS

```
Error in Predicting Latitude:
Mean Absolute Error: 6.612467982780419e-05
Mean Squared Error: 5.857772308461443e-09
Root Mean Squared Error: 7.653608500871627e-05
Maximum error: 0.00014949614640613618
--- 1070.9163303375244 seconds ---
```

LATITUDE MODEL

```
Error in Predicting Longitude:
Mean Absolute Error: 0.00010863203350197637
Mean Squared Error: 1.568777933146181e-08
Root Mean Squared Error: 0.00012525086559166669
Maximum error: 0.00021503229709196603
--- 1070.9327092170715 seconds ---
```

LONGITUDE MODEL

PERFORMANCE OF MODEL

```
Predicting Location:
Error of unknown loc  1 :  4.938806277573051
Error of unknown loc  2 :  17.64410609885038
Error of unknown loc  3 :  14.56878511151492
Error of unknown loc  4 :  11.265516993581313
Error of unknown loc  5 :  16.75676251593451
Error of unknown loc  6 :  9.345430964298226
Error of unknown loc  7 :  13.587220015745304
Error of unknown loc  8 :  16.462637949794836
Error of unknown loc  9 :  9.462216708598106
Error of unknown loc  10 :  17.54461631055337
Error of unknown loc  11 :  17.786571000997192
Error of unknown loc  12 :  18.320246462133195
Error of unknown loc  13 :  7.653023663509865
```

```
Maximum error in metres :  18.320246462133195
Average error in metres :  13.487380005621866
```

EVALUATION METRICS

```
Maximum error in metres :  18.320246462133195
Average error in metres :  13.487380005621866
```

MAXIMUM AND AVERAGE ERROR

```
Latency
--- 0.00021123886108398438 seconds ---
```

COMPLEXITY OF THE MODEL

- The maximum error of estimation, also called the margin of error, is an indicator of the precision of an estimate.
- The average error of estimation, is an indicator of how close a regression line is to a set of points and gives the average for the set of errors. It acts as the reference error value giving us the maximum and minimum error that can be obtained from the model.
- Latency is a measurement that is used to determine the performance of the model. Latency refers to the time taken to process one unit of data provided only one unit of data is processed at a time. Hence it is an indicator of the complexity of the model.

CONCLUSION

CONCLUSION

- On the given dataset, we have implemented the AI/ML model called **Extra Trees regression model**, to localise the unknown node in combination with some data-preprocessing techniques.
- The techniques included, data restructuring, data smoothening, feature selection and data augmentation.
- Used 2 models, one to predict the latitude and one for the longitude.
- Obtained a maximum error of 18.3m with an average error of 13.48m in a 35x15m field.

REFERENCES

1. S. Hara and D. Anzai, "Experimental Performance Comparison of RSSI- and TDOA-Based Location Estimation Methods," in *VTC Spring 2008 - IEEE Vehicular Technology Conference*, May 2008, pp. 2651–2655. doi: 10.1109/VETECS.2008.581.
2. J. Xu, J. He, Y. Zhang, F. Xu, and F. Cai, "A Distance-Based Maximum Likelihood Estimation Method for Sensor Localization in Wireless Sensor Networks," *International Journal of Distributed Sensor Networks*, vol. 2016, pp. 1–8, Apr. 2016, doi: 10.1155/2016/2080536.
3. A. Poulose and D. S. Han, "Hybrid Deep Learning Model Based Indoor Positioning Using Wi-Fi RSSI Heat Maps for Autonomous Applications," Electronics, vol. 10, no. 1, Art. no. 1, Jan. 2021, doi: 10.3390/electronics10010002.
4. R. Mino *et al.*, "A Belief Propagation-Based Iterative Location Estimation Method for Wireless Sensor Networks," in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5. doi: 10.1109/PIMRC.2006.254179.
5. R. Zemek, S. Hara, K. Yanagihara, and K. Kitayama, A Joint Estimation of Target Location and Channel Model Parameters in an IEEE 802.15.4-based Wireless Sensor Network. 2007, p. 5. doi: 10.1109/PIMRC.2007.4394355.
6. R. Zemek, D. Anzai, S. Hara, K. Yanagihara, and K. Kitayama, "RSSI-based Localization without a Prior Knowledge of Channel Model Parameters," *Int J Wireless Inf Networks*, vol. 15, no. 3, pp. 128–136, Dec. 2008, doi: 10.1007/s10776-008-0085-6.
7. R. Mino *et al.*, "A Novel Iterative Technique for Collaborative Location Estimations," in *2006 3rd International Symposium on Wireless Communication Systems*, Sep. 2006, pp. 564–568. doi: 10.1109/ISWCS.2006.4362362.

REFERENCES

8. R. Mino *et al.*, “A Belief Propagation-Based Iterative Location Estimation Method for Wireless Sensor Networks,” in *2006 IEEE 17th International Symposium on Personal, Indoor and Mobile Radio Communications*, Sep. 2006, pp. 1–5. doi: 10.1109/PIMRC.2006.254179.
9. N. Patwari, A. O. Hero, M. Perkins, N. S. Correal, and R. J. O’Dea, “Relative location estimation in wireless sensor networks,” *IEEE Transactions on Signal Processing*, vol. 51, no. 8, pp. 2137–2148, Aug. 2003, doi: 10.1109/TSP.2003.814469.

8.

10. Z. Khokhar and M. Siddiqi, “Machine Learning Based Indoor Localization using Wi-Fi and Smartphone,” Journal of Independent Studies and Research Computing - SZABIST, Vol: 18 – Issue: 1, Apr. 2021.
11. S. Hara and D. Anzai, *Comparison of Three Estimation Methods for RSSI-Based Localization with Multiple Transmit Antennas*. 2007, p. 3. doi: 10.1109/MOBHOC.2007.4428599.
12. D. Anzai and S. Hara, “An RSSI-Based MAP Localization Method with Channel Parameters Estimation in Wireless Sensor Networks,” in *VTC Spring 2009 - IEEE 69th Vehicular Technology Conference*, Apr. 2009, pp. 1–5. doi: 10.1109/VETECS.2009.5073388.
13. S. Hara, D. Anzai, T. Yabu, K. Lee, T. Derham, and R. Zemek, “A Perturbation Analysis on the Performance of TOA and TDOA Localization in Mixed LOS/NLOS Environments,” *IEEE Transactions on Communications*, vol. 61, no. 2, pp. 679–689, Feb. 2013, doi: 10.1109/TCOMM.2013.012313.110509.

THANK YOU