Predictive Wi-Fi Network Selection: Enhancing User Experience through an ML-based Approach

Akash V  
*dept. name of organization   
(of Affiliation)*  
*name of organization   
(of Affiliation)*City, Country  
email address or ORCID

Manoj Howale S   
*dept. name of organization   
(of Affiliation)*  
*name of organization   
(of Affiliation)*City, Country  
email address or ORCID

Ajit J Gupta   
*dept. name of organization   
(of Affiliation)*  
*name of organization   
(of Affiliation)*City, Country  
email address or ORC ID

Given Name Surname   
*dept. name of organization   
(of Affiliation)*  
*name of organization   
(of Affiliation)*City, Country  
email address or ORC ID

Given Name Surname   
*dept. name of organization   
(of Affiliation)*  
*name of organization   
(of Affiliation)*City, Country  
email address or ORC ID

*Abstract*—With the increasing use of Wi-Fi in our devices at home and in offices, etc. it's imperative to connect to an AP which provides seamless and smooth performance. This research paper addresses the need for an efficient solution to predict and select the most suitable AP, thus enhancing the user's quality of application usage. We have worked on a ML model which takes care of different parameters such as signal strength, bandwidth, distance, frequency, traffic parameters and previous connections with AP and predicts the best AP user should connect to. This research which has been proposed by this paper could potentially revolutionize the way transitions are made between networks to make the process smooth and latency free.

Keywords—WiFi, Best AP, Machine Learning

# INTRODUCTION

With the increase in demand of the internet today all cell phones have 4G/5G cellular data network, but this also is not sufficient for the user's needs. Users are adopting Wi-Fi technology at a faster pace than before to use home, offices and public spaces. It has become an integral part of our lives. However, there are several problems faced by users while connecting to the Wi-Fi. Sometimes it becomes difficult for the user to choose which AP to select among the available AP. It's because with just the signal strength shown in the device it's hard to predict which will be the best AP. Once you are connected to the AP there will be moments when you have a better option to connect to some other AP then your present one. At this point our research work comes needy for the user. it has become imperative to establish a seamless and reliable connection to an Access Point (AP) that can provide optimal performance. This research paper aims to address the pressing need for an efficient solution to predict and select the most suitable AP, thereby enhancing the overall user experience and quality of application usage.

In this paper, we have developed a machine learning (ML) model that considers various crucial parameters in AP selection, including signal strength, bandwidth availability, distance, frequency, traffic parameters, and the user's previous connections with APs.

The work is focused on the parameters which are accessible by the android API. By leveraging these parameters, our model is capable of predicting the best APP for users to connect to, thereby streamlining the process of network transition and ensuring minimal latency.

The implications of this research are far-reaching and have the potential to revolutionize the way users seamlessly transition between networks. By employing our ML model, users can expect smoother and latency-free transitions, allowing for uninterrupted and high-quality application usage.

The major objective is to create a model which can find the most appropriate network which is suitable for the use case of the customer, with least traffic and high RSSI value.

Ultimately, the findings presented in this research paper contribute to the advancement of network connectivity and usability, benefiting users in both domestic and professional environments. The ability to accurately predict and select the most suitable AP promises to enhance the overall efficiency and performance of Wi-Fi networks, thereby elevating the user's quality of experience in an increasingly connected world.

# LITERATURE SURVEY

This literature survey aims to study and learn more about the various approaches and parameters that were used to find the best access point. Researchers have made some related research on various methods that

Muhammad Asif Khan (2020) proposed a method to calculate transmission throughput of real time Wi-Fi networks using several ML models such as multilayer perceptron (MLP), support vector regressors (SVR), decision trees (DT) and random forests (RF). The author used no. stations connected to AP, signal strength at each station, modulation scheme, data rates, inter-arrival time, packet arrival rate, channel parameter and number of retransmissions as parameters in this paper.

Phillip B. Oni and Steven D. (2016) proposed a decentralized AP selection scheme that takes interference at the candidate APs into account and selects AP that offers best signal-interference-plus noise ratio (SINR). The model used RSS, Probe request and probe response as parameters and applied the algorithm DL-SINR AP Selection Algorithm and Optimal AP Selection Algorithm (OPASA).

Biljana Bojović (2011) proposed a cognitive AP selection scheme based on a supervised learning approach which will enable the mobile station to select the AP that offers the best performance. This paper used signal to noise ratio, probability of failure, business ratio, average beacon delay, number of detected stations as parameters and applied it to the Multi-layer Feed-forward Neural Network (MFNN) model.

Vismika Ranasinghe (2021) proposed a graph neural network (GNN) based access point (AP) selection algorithm for cell-free massive multiple-input multiple-output (MIMO) systems using reference signal received power (RSRP) and distance between devices to all AP’s and distance between AP to all other access points.

Payam Porkar Rezaeiye (2021) proposed choosing the access point by Markov game to enhance the load balancing and efficiency in IoT networks based on Wi-Fi and Li-Fi combination. The paper used Transmission rate and delay ratio as parameters.

In, Muhammad Asif Khan (2022) proposed a data-driven machine learning (ML) schemes to efficiently solve these problems in wireless LAN (WLAN) networks. The paper was implemented on Cognitive Wi‑Fi Networks using the parameters Clients, RSSI, Noise level, MAC queue and Time stamp on following models’ random forest, Multi-layer Perceptron (MLP) and SVR.

In, Davi Militani implemented a Resource Allocation service for wireless networks using Random Forest machine learning algorithm. The RF algorithm applied to heterogeneous network technologies to determine the AP selection strategy by using 12 network parameters.

In, Zayan El Khaled (2022) proposed two algorithms that predict the success of a user association to FWN via the analysis of user feedback and subscription status, in combination with radio frequency parameters. The algorithms used are Nearest Neighbour and Deep Nearest neighbour.

In, Pengyu Huang (2020) proposed a indoor localization algorithm MAPS (indoor localization algorithm based on multiple access point selection). The author considered the parameters RSSI, CSI (Channel State Information) using the K-means algorithm.

In, S. Vasudevan (2005) proposed a methodology for the estimation of potential upstream and downstream bandwidth between a client and an AP based on measurements of delay incurred by 802.11 Beacon frames from the AP.

In, Xiaohuan Yan (2008) proposed a handover decision method based on the prediction of traveling distance within an IEEE 802.11 wireless local area network (WLAN) cell by using the parameter RSS change rate.

Alessandro Raschellà (2019) proposed an access Point (AP) allocation algorithm for dense Wi-Fi networks, which relies on a centralized potential game developed in a Software-Defined Wireless Networking (SDWN)-based framework. The proposed strategy optimizes the allocation of the Wi-Fi stations (STAs) to APs and allows their dynamic reallocation according to possible changes in the capacity of the Wi-Fi network.

In, Sunghwan Kim (2020) worked on Access point selection in a n cell-free massive multiple-input multiple-output (MIMO) systems, where APs equipped with a large number of antennas are geographically distributed over a wide area with no cell border.

# Methodology

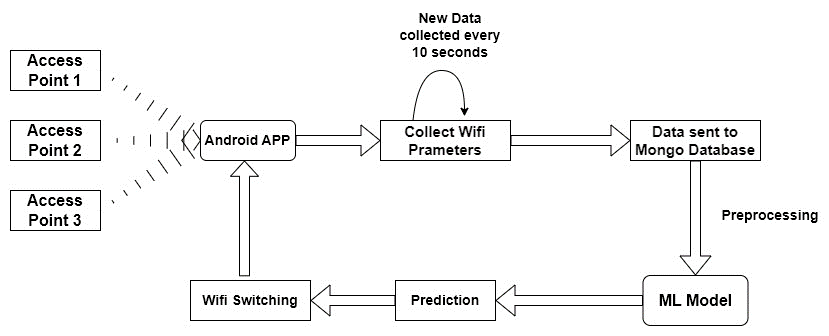


Figure 1: Data Flow Diagram

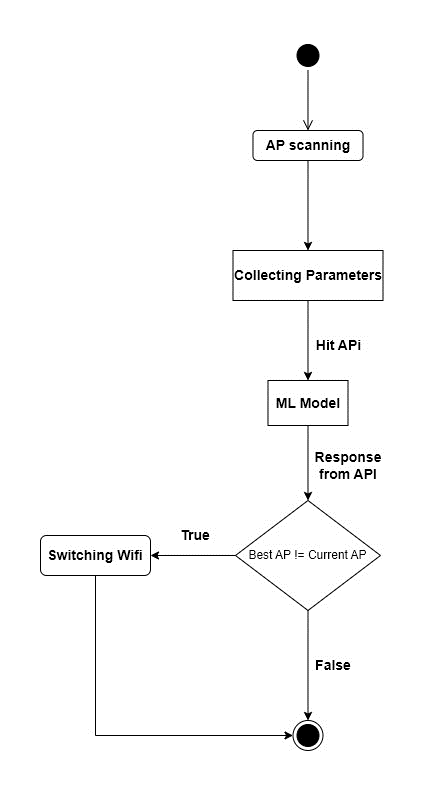


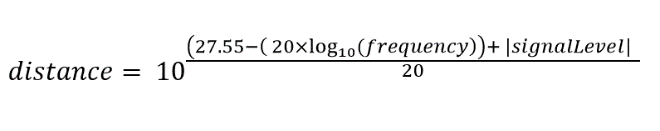
Figure 2: Control Flow Diagram

As shown in Figure1 and Figure2 our Android application will scan the nearby APs using the built in methods of Android API. We will extract the relevant parameters using Android API and these parameters will be sent to a cloud database i.e. MongoDB Atlas. The data stored in cloud is used to train the model. Once the model is trained we have deployed using FlaskAPI. Now the scanned parameters of the Aps are sent to our model using the API and a response is returned which is the predicted score of that AP. Based on the predicted score of all the APs the user can connect the AP with the maximum score at that time.

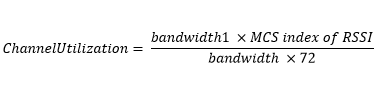
### Data Collection: We have created an Android app which collects real time dataset with complete variance, as our application remains running in the backgroung while the user is using the device normally. We gathered different wifi parameters such as SSID, RSSI, Frequency, Channel Bandwidth, Link Speed, Security, RxBytes, TxBytes, RxPackets, TxPackets, number of times AP was previously connected, Distance from Access Point, Channel Utilisation and Level. These parameters are collected from each access point available in the vicinity of the user device. The collected parameters are sent to a cloud database, which we have implemented using MongoDB Atlas. The android app scans every 10 seconds to capture all variations in the wifi network.

### Parameters Collected: The distinct parameters used for creating the model are elucidated in the below paragraphs:

* **RSSI** : The RSSI is a crucial parameter used in wireless networks to measure the strength of the received signal from an access point (AP). By capturing the signal strength, the model can determine the quality of the connection and make informed decisions about the best access point to connect to. The RSSI provides valuable insights into the signal propagation characteristics, allowing the model to prioritize APs with stronger signals for enhanced network performance and stability.
* **Distance from access point** : The proximity to an access point plays a significant role in determining the quality of the wireless connection.. Machine learning models can utilize distance as a feature to optimize access point selection. By incorporating this information, the model can prioritize APs that are closer to the user device, minimizing signal attenuation and potential interference.



* **Channel utilisation** : Channel utilisation refers to the level of activity or congestion on a specific channel within a wireless network. It is crucial to select an access point operating on an uncongested channel to minimize interference and maximize network performance.  By considering channel utilization as a feature, the model can select an AP operating on a less crowded channel, resulting in improved throughput and reduced latency.



* **Previously connected networks :** Understanding the user's historical connections to different access points can provide valuable insights for access point selection. By analyzing previously connected networks, machine learning models can learn user preferences and tendencies. By leveraging this information, the model can prioritize access points that have shown better performance in previous connections, enhancing the user experience and ensuring a seamless transition between APs.

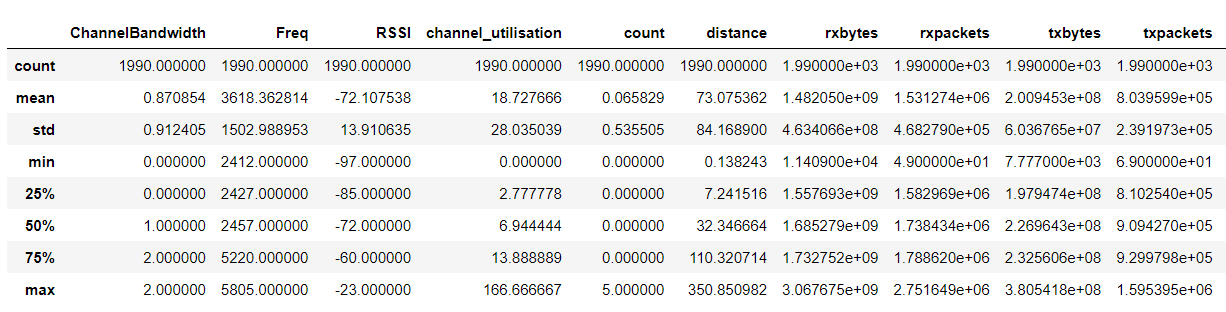


Figure 3: Data description

* **Traffic Parameters:** These metrics, TxBytes, TxPackets, RxBytes, and RxPackets, are commonly used for network monitoring, troubleshooting, and performance analysis. By tracking and analyzing these metrics, network administrators can gain insights into data transmission and reception patterns, identify potential issues or bottlenecks, and optimize network performance accordingly.

Score from these parameters is calculated as

The final Score is :

Final Score = Nscore + trafficScore

### Data Visualization:

The data is collected with all the possibilities of variance in RSSI level, distance of AP and other parameters. Same can be depicted by the figures shown below:

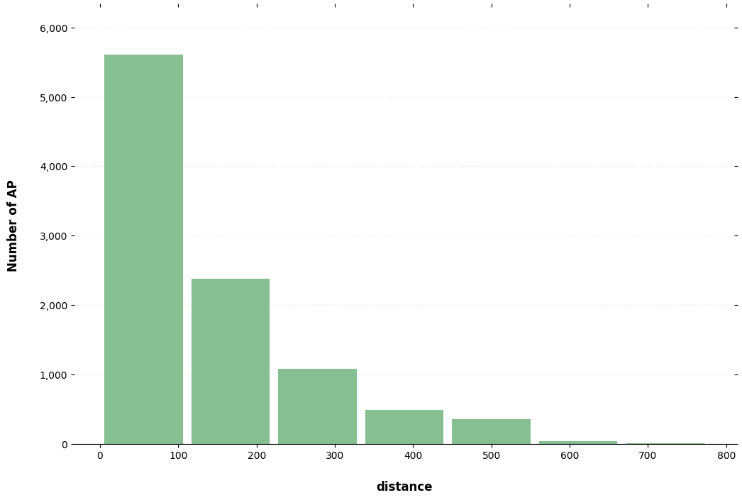


Figure 4: Distance vs Number of AP

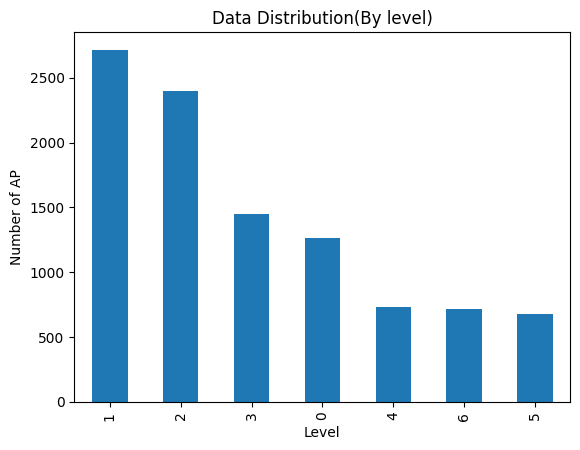


Figure 5: Level vs Number of AP

### Data preprocessing:

We conducted a thorough analysis of our dataset using various plotting methods and carefully examined the presence of NULL values. The data was collected periodically, every 10 from an Android device using the android.wifi.scan() function to prevent redundancy. However, we encountered a challenge when the user remained stationary with the application running, resulting in multiple data points with minimal variation. Therefore, we only considered data points where the user was relatively mobile, allowing us to obtain a more diverse and informative dataset. After obtaining the required rows from cloud database, we proceeded to drop columns that were not deemed significant features. Subsequently, we plotted additional graphs to gain further insights into the dataset. We addressed the issue of outliers by employing the methods to remove them and we to balance the dataset as much as possible.

For testing and training purposes, we divided the dataset in a 20:80 ratio, allocating 20% for testing and 80% for training.

### Machine learning models :

**Linear Regression:** Linear regression is a widely used statistical technique for modelling the relationship between a dependent variable and independent variables. It aims to estimate the parameters of a linear equation that best describes the linear association. With its simplicity and interpretability, linear regression provides insights into the relationship strength and direction, enabling prediction, hypothesis testing, and understanding variable impact. In this research paper, we employ linear regression to analyse the relationship between our dependent variable and independent variables. The findings contribute to understanding the phenomenon and inform decision-making in relevant fields.

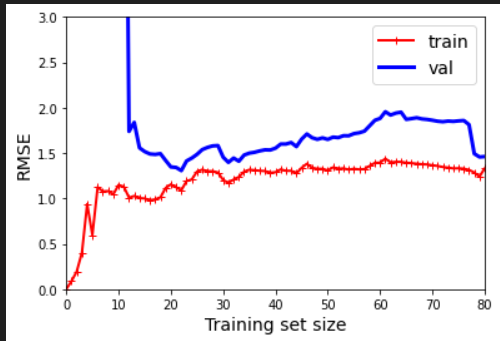


Figure 6: RMSE error

**Decision Tree Regressor:**

Decision tree regression is a versatile machine learning approach used for predicting continuous numerical values. It builds a tree-like model where each internal node represents a feature condition, and each leaf node corresponds to a predicted value. With interpretability, non-linear relationship modeling, and robustness to outliers, decision tree regression is a valuable tool. In this research paper, we utilize decision tree regression to analyze the relationship between our independent variables and the target variable. By uncovering patterns and making accurate predictions, our study contributes to understanding the underlying phenomenon and its practical implications in various domains**.**

Mean squared error: 0.178

R-squared: 0.999

Accuracy: 99.45 %.

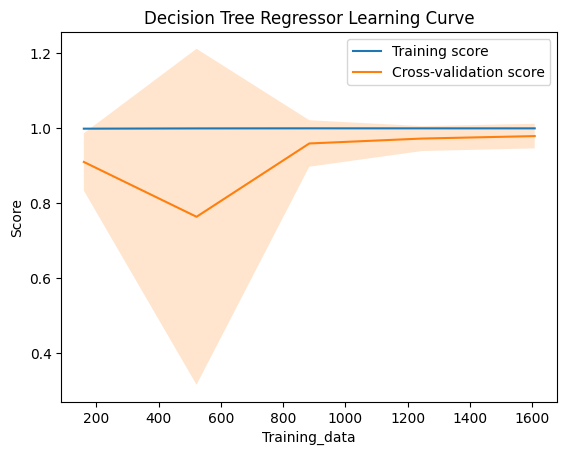
****

Figure 7:Learning Curve

**Random Forest Regressor :** This method has gained widespread recognition as a powerful ensemble learning method. By leveraging the concept of decision trees and incorporating randomness through random feature selection and bootstrap aggregating, Random Forest overcomes the limitations of individual decision trees, such as overfitting and high variance. It combines predictions from multiple trees to improve accuracy, handle outliers and missing data robustly, estimate feature importance, and capture complex non-linear relationships. With its efficiency in handling large datasets and its ability to provide interpretable results, Random Forest was chosen as one of the models.

Mean squared error: 0.329

R-squared: 0.998

Accuracy: 97.41 %.

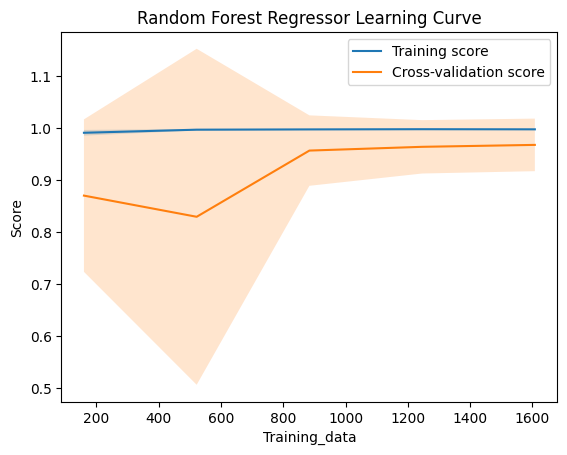


Figure 8: Learning Curve

ALGORITHM

1.

2. Loop continuously for real-time parameter collection:

a

3. Check dataset variability:

a. If variance is below a threshold:

i. Continue collecting parameters until desired variance is met.

4. Dataset variability is sufficient:

a. Append collected parameters to a CSV file, creating a new row for each entry. . Collect parameters (e.g., signal strength, bandwidth, latency) from available APs.

b. Store parameters in a temporary data structure or dataframe.

5. Data preprocessing:

a. Handle missing values, normalize or scale data, and encode categorical variables if necessary.

6. Split dataset:

a. Divide the dataset into training and testing sets using a suitable ratio (e.g., 80% for training, 20% for testing).

7. Model selection:

a. Choose appropriate machine learning algorithm(s) for best AP selection (e.g., decision tree, random forest, neural network).

8. Model training:

a. Train selected machine learning model(s) using the training dataset.

9. Model evaluation:

a. Evaluate performance of the trained model(s) using suitable evaluation metrics (e.g., accuracy, precision, recall, F1-score) on the testing dataset.

10. Optional hyperparameter tuning:

a. Perform hyperparameter tuning on the model(s) to optimize performance.

11. Save the trained model(s) to disk for future use.

12. Continuously monitor real-time parameters:

a. When new parameters are received:

i. Preprocess the parameters.

ii. Use the trained model(s) to predict the best AP selection.

13. Repeat steps 2-12 as required for continuous updating and refinement of the model.

# Conclusion

In this paper we have proposed a ML based Wi-Fi score prediction in a mobile device. Our method collects Wi-Fi parameters from android devices and using these parameters we generate a score from our ML model to calculate the scores. This method also takes into consideration of past connected Wi-Fi networks and other parameters such as channel utilization in the ML model for score prediction. All tests and data collection were performed on IEEE 802.11 based technology. The score predicted by the ML model show the Access Point with the best performance even in a highly varying environment.

### 

##### Acknowledgments

“Acknowledgment(s)” is spelled without an “e” after the “g” in American English.

As you can see, the formatting ensures that the text ends in two equal-sized columns rather than only displaying one column on the last page.

This template was adapted from those provided by the IEEE on their own website.

##### References

1. D. V. Lindberg and H. K. H. Lee, “Optimization under constraints by applying an asymmetric entropy measure,” *J. Comput. Graph. Statist.*, vol. 24, no. 2, pp. 379–393, Jun. 2015, doi: 10.1080/10618600.2014.901225.
2. B. Rieder, *Engines of Order: A Mechanology of Algorithmic Techniques*. Amsterdam, Netherlands: Amsterdam Univ. Press, 2020.
3. I. Boglaev, “A numerical method for solving nonlinear integro-differential equations of Fredholm type,” *J. Comput. Math.*, vol. 34, no. 3, pp. 262–284, May 2016, doi: 10.4208/jcm.1512-m2015-0241.
4. Muhammad Asif Khan, Ridha Hamila, Nasser Ahmed Al-Emadi, Serkan Kiranyaz, Moncef Gabbouj, Real-time throughput prediction for cognitive Wi-Fi networks, Journal of Network and Computer Applications, Volume 150, 2020, 102499, ISSN 1084-8045, <https://doi.org/10.1016/j.jnca.2019.102499>.
5. P. B. Oni and S. D. Blostein, "Decentralized AP selection in large-scale wireless LANs considering multi-AP interference," 2017 International Conference on Computing, Networking and Communications (ICNC), Silicon Valley, CA, USA, 2017, pp. 13-18, doi: 10.1109/ICCNC.2017.7876094.
6. B. Bojovic, N. Baldo, J. Nin-Guerrero and P. Dini, "A supervised learning approach to cognitive access point selection," 2011 IEEE GLOBECOM Workshops (GC Wkshps), Houston, TX, USA, 2011, pp. 1100-1105, doi: 10.1109/GLOCOMW.2011.6162348.
7. Muhammad Asif Khan, Ridha Hamila, Nasser Ahmed Al-Emadi, Serkan Kiranyaz, Moncef Gabbouj, Real-time throughput prediction for cognitive Wi-Fi networks, Journal of Network and Computer Applications, Volume 150, 2020, 102499, ISSN 1084-8045, <https://doi.org/10.1016/j.jnca.2019.102499>.
8. Porkar Rezaeiye, P., Sharifi, A., Rahmani, A.M. et al. Access point selection in the network of Internet of things (IoT) considering the strategic behavior of the things and users. J Supercomput 77, 14207–14229 (2021). <https://doi.org/10.1007/s11227-021-03788-3>
9. Khan, M.A., Hamila, R., Gastli, A. et al. ML-Based Handover Prediction and AP Selection in Cognitive Wi-Fi Networks. J Netw Syst Manage 30, 72 (2022). <https://doi.org/10.1007/s10922-022-09684-2>
10. D. Militani, S. Vieira, E. Valadão, K. Neles, R. Rosa and D. Z. Rodríguez, "A Machine Learning Model to Resource Allocation Service for Access Point on Wireless Network," 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2019, pp. 1-6, doi: 10.23919/SOFTCOM.2019.8903853.
11. D. Militani, S. Vieira, E. Valadão, K. Neles, R. Rosa and D. Z. Rodríguez, "A Machine Learning Model to Resource Allocation Service for Access Point on Wireless Network," 2019 International Conference on Software, Telecommunications and Computer Networks (SoftCOM), Split, Croatia, 2019, pp. 1-6, doi: 10.23919/SOFTCOM.2019.8903853.
12. Huang, P., Zhao, H., Liu, W. et al. MAPS: Indoor Localization Algorithm Based on Multiple AP Selection. Mobile Netw Appl 26, 649–656 (2021). <https://doi.org/10.1007/s11036-019-01411-7>
14. X. Yan, N. Mani and Y. A. Sekercioglu, "A Traveling Distance Prediction Based Method to Minimize Unnecessary Handovers from Cellular Networks to WLANs," in IEEE Communications Letters, vol. 12, no. 1, pp. 14-16, January 2008, doi: 10.1109/LCOMM.2008.071430.
15. Alessandro Raschellà, Faycal Bouhafs, Michael Mackay, Qi Shi, Jorge Ortín, José Ramón Gállego, Maria Canales, A dynamic access point allocation algorithm for dense wireless LANs using potential game, Computer Networks, Volume 167, 2020, 106991, ISSN 1389-1286, <https://doi.org/10.1016/j.comnet.2019.106991>.
16. S. Biswas and P. Vijayakumar, "AP selection in Cell-Free Massive MIMO system using Machine Learning Algorithm," 2021 Sixth International Conference on Wireless Communications, Signal Processing and Networking (WiSPNET), Chennai, India, 2021, pp. 158-161, doi: 10.1109/WiSPNET51692.2021.9419450.