Contents

List of Figures

List of Tables

Predicting Indoor Positioning using Machine Learning

İsmail Emre Alci
Marmara University
Dept. of Electrical & Electronics Engineering
Dept. of Computer Engineering

1 Abstract

This project aims to leverage machine learning techniques to enhance the precision of indoor location prediction, especially in environments where Global Positioning System (GPS) signals are attenuated. The focus lies on addressing the challenges posed by limited GPS visibility and fluctuating signal strengths in indoor spaces. By harnessing data from wireless local area networks (WLAN) and other indoor positioning technologies, this study seeks to develop predictive models capable of accurately determining indoor locations. The dataset "trainingData.csv" contains a variety of attributes, including WLAN signal measurements, geographic coordinates, floor levels, and user-related information. To streamline analysis, Principal Component Analysis (PCA) is implemented to condense the dimensionality of WLAN data while retaining its inherent patterns. Subsequently, three distinct machine learning algorithms—K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM)—are employed for predicting indoor locations. Furthermore, a Deep Learning approach using Artificial Neural Networks (ANN) is introduced, allowing for a comparative assessment of the effects of PCA within the Deep Learning framework. The study culminates in a comprehensive comparison of accuracy scores for various algorithms, both with and without PCA, shedding light on the intricate interplay between dimensionality reduction and indoor location prediction.

2 Introduction

In the realm of location-based services, the distinction between outdoor and indoor environments holds paramount importance. While outdoor spaces are commonly facilitated by the accuracy of the Global Positioning System (GPS) for location determination, indoor areas often suffer from weakened GPS signals, presenting challenges for precise indoor localization. This project centers around harnessing the capabilities of machine learning to overcome these challenges and enhance indoor localization accuracy.

With a specific focus on indoor environments, this project addresses the complexities stemming from limited GPS visibility and fluctuating signal strengths indoors. By integrating machine learning algorithms, the aim is to construct prediction models capable of accurately determining indoor locations based on available data. This endeavor holds significant potential for applications such as navigation, resource management, and enriching user experiences within indoor spaces.

The project's approach involves the fusion of data from wireless local area networks (WLAN), including Wi-Fi signals and other indoor positioning technologies. By generating and training machine learning models, the objective is to predict indoor locations with heightened precision and efficiency. The subsequent sections delve into the specifics of data preprocessing, the application of various machine learning algorithms, the incorporation of Deep Learning techniques, and a comparative analysis of their results. This comprehensive approach seeks to shed light on the intricate relationship between machine learning, dimensionality reduction, and accurate indoor location prediction.

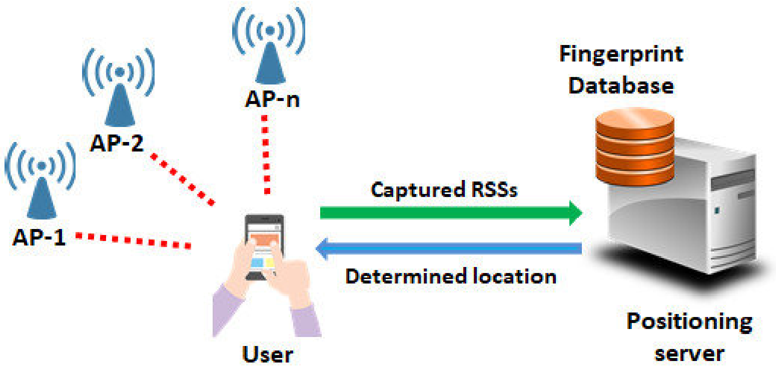

Figure 1: Illustration of Wi-Fi Fingerprint-based IPS

3 Data Overview and Preprocessing

The dataset, named "trainingData.csv," comprises 19937 rows and 529 columns. Notably, the first 520 columns represent distinct WAP signals, ranging from WAP01 to WAP520. The remaining columns correspond to essential attributes including *Longitude*, *Latitude*, *Floor*, *BuildingID*, *SpaceID*, *RelativePosition*, *UserID*, *PhoneID*, *TimeStamp*.

Given the substantial size of the dataset, we initiate a preprocessing step to enhance analysis efficiency. To achieve this, we employ Principal Component Analysis (PCA), a mathematical technique used to reduce the dimensionality of the data while retaining its essential information. In particular, the 520 WAP columns (independent variables) undergo PCA transformation into a compact set of 100 PCA columns. This process enables us to capture the most salient features of the WAP data.

It's crucial to note that while PCA is applied to the x (independent) variables, the y variables (dependent variables) remain unaffected. This ensures the integrity of the labels associated with the data.

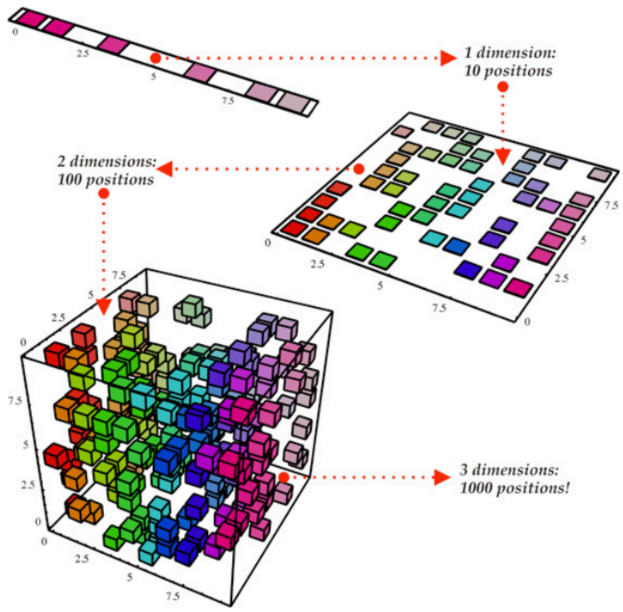

Figure 2: Overview of Principal Component Analysis

4 Machine Learning Application

Following the application of PCA, we consolidate the PCA-transformed x variables with the unaltered y variables and other pertinent attributes. This amalgamated dataset lays the groundwork for the implementation of machine learning algorithms. Our primary objective is to enhance the precision of predicting indoor locations. We leverage the prowess of three distinct machine learning algorithms: K-Nearest Neighbors (KNN), Random Forest, and Support Vector Machine (SVM). These algorithms are well-suited for classification tasks and are poised to facilitate the accurate prediction of indoor locations, capitalizing on the strength of the consolidated dataset. Subsequent sections delve deeply into the methodology and intricacies of each machine learning algorithm. By synergizing the benefits of dimensionality reduction and advanced machine learning techniques, our aspiration is to elevate the precision and reliability of our indoor location predictions.

5 Deep Learning Application

In this phase, we will emulate the Machine Learning approach discussed earlier by employing Deep Learning and Artificial Neural Networks (ANN). Similar to our Machine Learning methodology, we will initially determine accuracy values without implementing PCA. Subsequently, we will apply PCA and re-evaluate the accuracy values. This systematic approach allows us to directly compare these outcomes with the results obtained from the KNN, Random Forest, and SVM algorithms in the Machine Learning phase.

By replicating our project with Deep Learning, we seek to ascertain the impact of PCA within this framework, enhancing our understanding of how dimensionality reduction affects Deep Learning performance. This comparative analysis aims to provide valuable insights into the interplay between dimensionality reduction techniques and the intricacies of different learning algorithms.

6 Machine Learning vs Deep Learning

Machine learning involves the development of algorithms that enable computers to learn patterns from data and improve their performance on specific tasks through experience, often relying on engineered features. Deep learning is a subset of machine learning that focuses on training deep neural networks to automatically learn intricate features and representations from raw data, eliminating the need for manual feature engineering. It excels in tasks like image and speech recognition, language processing, and more complex pattern recognition.

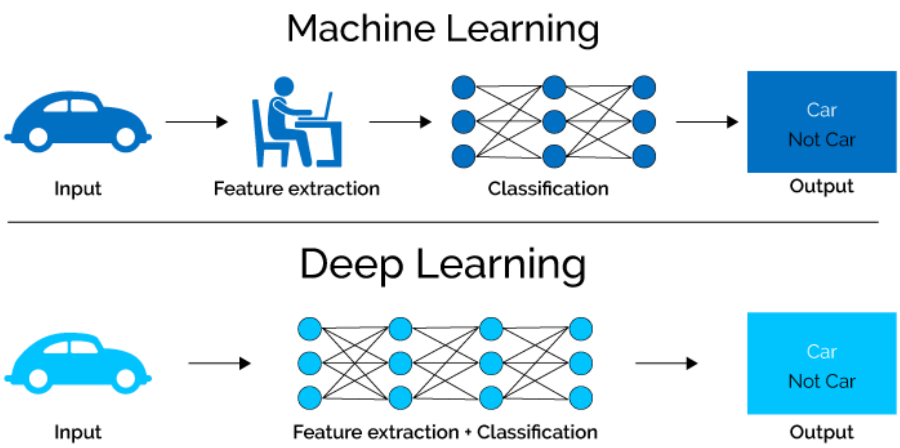

Figure 3: Difference between Machine Learning and Deep Learning

7 Results and Table

We have calculated accuracy scores for each algorithm (KNN, Random Forest, SVM, ANN) under both PCA-transformed and non-transformed conditions. These results are presented in a structured table format, allowing a direct comparison of predictive performances for indoor locations across diverse attributes. We will analyze and compare these values in the Conclusion section.

	ML						**DL**	
	KNN		**Random Forest**		**SVM**		**ANN**	
	PCA	*No-PCA*	*PCA*	*No-PCA*	*PCA*	*No-PCA*	*PCA*	*No-PCA*
Target 3-FLOOR	96.288	97.0411	97.0160	99.5737	94.4583	97.3671	96.9157	93.2296
Target 4-BUILD ID	99.8495	99.7241	99.7743	99.7993	99.6740	99.7993	99.87462	99.7492
Target 6-REL. POSITION	93.4052	93.1795	93.5305	95.7873	87.5125	91.0481	93.3550	87.4122

Table 1: Accuracy Scores of Different Algorithms

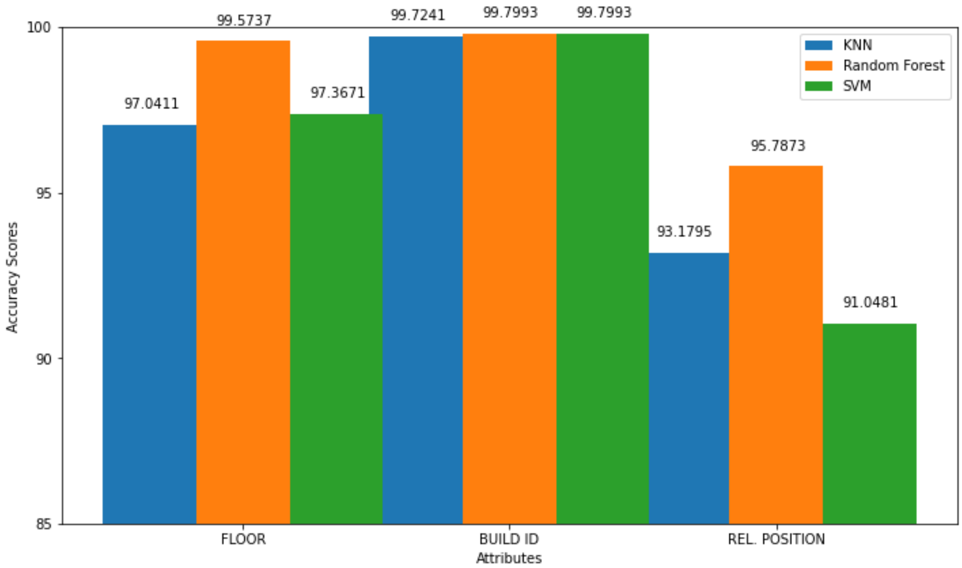

Figure 4: Comparison between ML scores for No-PCA

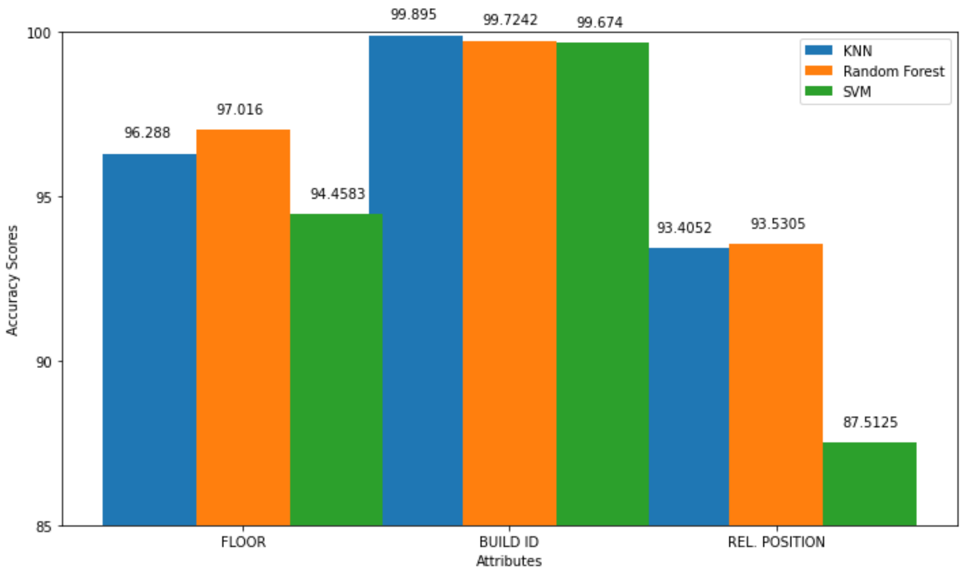

Figure 5: Comparison between ML scores for PCA

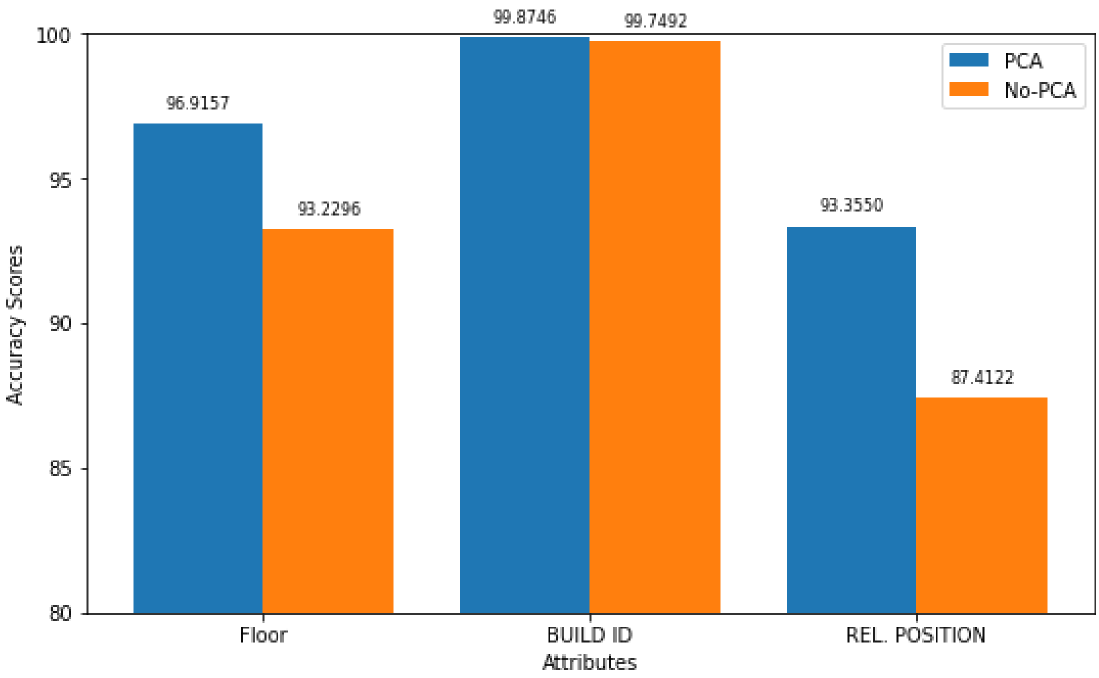

Figure 6: Accuracy Scores for ANN with and without PCA

8 Conclusion

Conclusively, after a thorough analysis of Table 1, it is evident that in our indoor location prediction project's Machine Learning phase, the SVM algorithm achieves the lowest accuracy score. On the other hand, looking at Figure 4 and Figure 5, as we will clearly observe, the Random Forest algorithm reaches the highest accuracy score However, it is worth noting that the accuracy scores of the KNN and Random Forest algorithms are quite close, nearly indistinguishable from each other. These findings highlight the varying suitability of different algorithms for distinct features and data structures, aiding us in evaluating the overall performance of our project.

In the context of Deep Learning, when delving into the creation of Artificial Neural Network (ANN) models, an intriguing observation emerges. As depicted in Figure 6, it becomes apparent that the inclusion of Principal Component Analysis (PCA) in the ANN model often leads to improved outcomes. Given that larger datasets usually confer advantages to ANN models, this result might appear somewhat counterintuitive. Nevertheless, it's important to note that the application of PCA to ANN doesn't necessarily guarantee poor outcomes. As exemplified in this case, using PCA in conjunction with ANN can indeed yield enhanced results

As a result, applying PCA does not always lead to an increase in accuracy score. The impact of PCA on accuracy depends on several factors:

- **Data Structure:** The nature and distribution of the dataset can influence the effect of PCA. If the data is linearly separable, PCA might have a positive impact on accuracy.
- **Distribution of Variance:** PCA aims to capture the highest variance. If the variance across different classes is significantly different, applying PCA might cause loss of information from classes with lower variance, potentially leading to a decrease in accuracy.
- **Dimension Reduction:** PCA reduces dimensions, but excessive reduction can result in information loss. If a significant portion of features is discarded, the reduction in complexity might lead to decreased accuracy.
- **Class Separability:** PCA's impact depends on how well the classes are separable. If classes are mixed or overlapping, PCA may not necessarily enhance accuracy.