Journal of Information Systems and Telecommunication
Vol.13, No.1, January-March 2025, 63-76
http://jist.acecr.org
ISSN 2322-1437 / EISSN:2345-2773

Enhancing IoT Device Behavior Prediction through Machine Learning Models

Shubham Minhass[1*], Ritu Chauhan[2] , Harleen Kaur[3]

[1].Amity Institute of Information Technology, Amity university, Noida, Uttar Pradesh, India
[2].Artificial Intelligence and IoT Lab, Center for Computational Biology and Bioinformatics, Amity university, Noida, Uttar Pradesh, India
[3].Department of Computer Science and Engineering, Jamia Hamdard, Delhi, India

Received: 06 Aug 2024/ Revised: 04 Feb 2025/ Accepted: 04 Mar 2025

Abstract

There is an urgent need for precise and trustworthy models to forecast device behavior and evaluate vulnerabilities as a result of the Internet of Things' (IoT) explosive growth. By assessing the effectiveness of several machine learning algorithms logistic regression, decision trees, random forests, Naïve Bayes, and KNN on two popular IoT devices Alexa and Google Home Mini this study seeks to enhance IoT device behavior forecasting. Our results show that Naïve Bayes and random forest models are more accurate and efficient than other algorithms at predicting device behavior. These findings demonstrate how important algorithm selection is for maximizing the performance of IoT systems. The study also emphasizes the usefulness of precise device behavior prediction for practical uses such as industrial control systems, home automation, and medical monitoring. For example, accurate forecasts can improve decision-making in crucial situations, facilitate more seamless automation, and stop system failures. In addition to adding to the expanding corpus of research on IoT data analysis, this study establishes the foundation for the creation of increasingly sophisticated machine learning models that can manage the intricate and ever-changing nature of IoT ecosystems. Future studies should concentrate on increasing the dataset's diversity to encompass a wider range of IoT environments and devices and enhancing the model's adaptability to changing IoT environments.

Keywords: Machine learning; predictive model; Smart Devices; Google Home Mini; Alexa; IoT; KNN.

1- Introduction

Numerous devices can now be connected to the internet thanks to the Internet of Things (IoT), which has increased their intelligence and efficiency. Some of the most popular examples of IoT technology are smart devices, like Google Home Mini and Amazon Alexa, which provide practical voice-activated features that make daily chores easier. However, one of the most important factors in determining user satisfaction and trust is how well these devices perform and comprehend voice commands. Accurate command recognition is particularly important in applications where these devices are used for sensitive tasks like financial services, home automation, or healthcare monitoring [1].

Although smart speakers are becoming more and more popular, little research has been done to assess how well they can recognize and react to voice commands. Customers rely on these devices to be extremely dependable, so any inaccuracy could irritate them or put sensitive applications at risk. Therefore, it is crucial to look into how accurate smart speakers are and how to use cutting-edge methods like machine learning to enhance their behavior prediction. The important question of how machine learning models can be applied to improve smart speaker prediction accuracy is the focus of this paper. The objective is to assess how well various machine learning algorithms predict the actions of gadgets like Google Home Mini and Alexa.

This study attempts to improve the performance of IoT devices by pointing out the advantages and disadvantages of different models. To do this, we extracted data from Alexa and Google Home Mini devices using Wireshark software. To analyze the data and evaluate how well these devices recognized different voice commands, a number of machine learning models were used, such as logistic regression, decision trees, random forest classifiers, and Naïve Bayes. The best machine learning algorithms for raising the predictive accuracy of IoT devices are identified by this analysis. This study's primary contribution is a

✉ **Shubham Minhass**
shubhamminhass@gmail.com

comparison of various machine learning algorithms and how they affect the accuracy of IoT devices [2].

Although earlier research has concentrated on the functionality of IoT devices, our work goes one step further by assessing these devices' predictive performance through the use of sophisticated machine learning techniques. The results have wider ramifications for enhancing IoT applications in practical
contexts where precision is essential, like home automation, healthcare, and industrial automation. Understanding the accuracy of IoT devices is essential for their effective utilization. It enables users to choose the right device for a specific task and ensures that the device performs the task correctly. Moreover, it helps developers to improve the accuracy of IoT devices, resulting in better user experience and satisfaction [3].
If IoT devices like voice assistants provide incorrect information due to low accuracy, it can have negative consequences for the user. For instance, if a user relies on a voice assistant to set an alarm to wake up in the morning, but the device fails to set the alarm accurately, the user may oversleep and be late for work or an important appointment. Similarly, if a user asks a voice assistant to play a specific song, but the device fails to recognize the command or plays the wrong song, it can lead to frustration and dissatisfaction. In some cases, inaccurate responses from IoT devices can lead to serious consequences, such as providing incorrect medical advice or inaccurate financial information. The accuracy of IoT devices is particularly critical in certain contexts, such as providing medical advice or financial information. If a user relies on a voice assistant to provide medical advice, inaccurate responses can have serious consequences such as misdiagnosis or recommending the wrong treatment. This can lead to severe health consequences for the user, including worsening of the medical condition or even death [4].
Similarly, if a voice assistant provides inaccurate financial information, it can lead to significant financial losses for the user. For example, if a user relies on a voice assistant to provide investment advice, and the device provides inaccurate information, it can lead to investment losses and financial instability. Inaccurate responses from IoT devices in these contexts can have severe consequences for the user. Therefore, it is crucial to ensure that the devices are accurate and reliable when providing such critical information. Developers of IoT devices must take into account the importance of accuracy in these contexts and implement necessary measures to ensure that the devices provide accurate and reliable information to users [5].
Moreover, if a user relies on a voice assistant for navigation purposes, and the device provides incorrect directions, it can lead to the user getting lost or ending up in the wrong location. This can be especially dangerous when driving, as it can lead to accidents and other related issues.
The significance of understanding the accuracy of IoT devices to ensure that they perform their intended functions correctly and reliably. It highlights that inaccurate responses from IoT devices can result in a range of consequences, ranging from minor inconveniences to serious and potentially dangerous situations. However, inaccurate responses from IoT devices can also have more severe consequences. For example, if a voice assistant provides incorrect directions, it can result in the user getting lost or ending up in the wrong location, which can be particularly dangerous when driving. Similarly, if a user relies on a voice assistant to set an alarm but the device fails to do so correctly, it can lead to the user being late for work or other important appointments [6]. Furthermore, the consequences of inaccurate responses can be particularly severe when it comes to medical advice or financial information. Inaccurate responses from a voice assistant in these contexts can lead to misdiagnosis, incorrect treatments, financial losses, and other serious consequences that can affect the user's health or financial stability.
However, it is essential to note that our research has some limitations. First, Since Google Home Mini and Alexa were the only two devices we looked into, it's possible that our conclusions don't apply to other smart speakers. Second, the accuracy of these devices in identifying other types of commands may vary, as our analysis is based on a restricted set of voice commands. Finally, our research is based on data collected after 2005, and previous studies conducted before this time may have different findings.
Not standing with these drawbacks, our study sheds light on smart speaker accuracy and emphasizes the significance of learning more about this feature of the products. Our research should help create smart speakers that are more dependable and accurate, which will enhance user satisfaction and spur more people to adopt IoT devices [7].

2- Objective

By using Wireshark software to record network traffic data and analysing it using different machine learning models, this research study investigates how well Alexa and Google Home Mini recognize and react to voice commands. The goal is to shed light on how accurate these smart speakers are. In order to classify and predict responses based on input features, the goals include using Wireshark to collect network traffic data, ensuring compliance with legal and ethical considerations while acknowledging potential limitations related to encrypted traffic; analysing the collected data using machine learning models like logistic regression, decision tree classifiers, random forest classifiers, and naïve Bayes models [8].
comparing model predictions with real device responses to evaluate how well the devices recognize different voice commands, such as playing music, sending reminders, and giving information; recognizing performance strengths and

shortcomings to suggest enhancements in speech recognition and natural language processing capabilities; and providing information that could help in the future development of more accurate and dependable devices, ultimately advancing smart speakers and improving IoT technology user experiences.we aim to provide insights into the performance of Google Home Mini and Alexa.

2-1- To Collect Data from Google Home Mini and Alexa Devices using Wireshark Software

Wireshark is a popular open-source packet sniffing and protocol analysis software that allows you to capture and analyze network traffic. To collect data from Google Home Mini and Alexa devices using Wireshark, you would need to connect your computer or laptop to the same network as the devices you want to monitor. This can typically be done by connecting to the same Wi-Fi network as the devices. Once you have connected to the network, you can start Wireshark and begin capturing packets. This involves selecting the appropriate network interface (e.g., Wi-Fi adapter) in Wireshark and starting a new capture session. It is worth noting that collecting data using Wireshark may have certain legal and ethical considerations, as it involves monitoring network traffic that may contain sensitive or private information. When using Wireshark to collect data from devices on a network, it is important to ensure that you have the necessary permissions and consents to do so. Without the right authorization, data collection can be illegal and unethical, with potentially dire repercussions including loss of trust and legal action. Therefore, it is important to obtain explicit consent from the owners of the devices being monitored, and to adhere to any applicable laws and regulations governing data privacy and security. Furthermore, the collected data may contain encrypted traffic that cannot be analyzed or decrypted using Wireshark alone. This can occur when the devices are communicating using encryption protocols such as SSL/TLS, which are designed to secure the communication and prevent eavesdropping. While it is possible to decrypt some types of encrypted traffic using Wireshark by capturing the necessary encryption keys, this can be a complex and time-consuming process, and may not be feasible in all cases. As a result, the collected data may be incomplete or limited in its usefulness for analysis purposes, particularly if the encrypted traffic contains important information related to the study being conducted. In order to guarantee the accuracy and completeness of the analysis, it is crucial to take into account any potential drawbacks of using Wireshark for data collection and, if needed, to supplement the data with information from other sources [9].

2-2- To Analyze the Collected Data using Different Machine learning Models, Including logistic Regression, Tree Classifier, Random Forest Classifier, and Naïve Bayes Model

The next stage is to use various machine learning models to analyse the data that has been gathered from Google Home Mini and Alexa devices. This entails constructing models that can correctly classify the data and generate predictions based on the input features by utilizing a variety of algorithms. One kind of linear regression that is used for classification tasks is the logistic regression model. Additionally, its ability to handle linearly separable data is one of its advantages, and it offers a clear understanding of how each input feature contributes to the final classification. Based on the values of the input features, the tree classifier model is a decision tree-based algorithm that divides the data into subsets recursively. It is also a flexible model for classification tasks because it can handle both continuous and categorical data. A probabilistic algorithm that presumes feature independence is the naïve Bayes model. Because of its capacity to manage sizable datasets and high-dimensional feature spaces, it is a well-liked option for text classification tasks like sentiment analysis, spam detection, and topic classification. In order to determine which class has the highest probability given the input features, Naive Bayes computes the probability of each class. This makes it computationally efficient and well-suited for large datasets. By using a combination of these machine learning models, it is possible to accurately analyze and classify the data collected from Google Home Mini and Alexa devices, providing valuable insights into their accuracy and performance [10].

2-3- To Determine the Accuracy of these Tevices in Recognizing and Responding to Different Types of Voice Commands, such as Playing Music, Setting Reminders, and Providing Information

The accuracy of smart speakers, such as Google Home Mini and Alexa, in recognizing and responding to different types of voice commands is a critical aspect of their overall performance. To determine this accuracy, the collected data can be used to train different machine learning models, as mentioned in the previous point. These models can then be used to predict the response of the smart speaker to a given voice command, based on the input features. Based on the features gleaned from the network data, the logistic regression model, for instance, can be used to forecast the likelihood that a smart speaker will correctly respond to a given command. Similar to this, based on the decision rules discovered from the data, the tree classifier and random forest classifier can be used to forecast the smart speaker's most likely response to a given command. By comparing the predictions of these models with the actual responses of the smart speaker to the same commands, we can determine the accuracy of the device in recognizing and responding to different types of voice commands. This information can be used to identify areas where the device may need improvement, such as in recognizing certain accents or understanding specific types of commands. Overall, determining the accuracy of smart speakers in recognizing and responding to voice commands is essential for evaluating their performance and identifying opportunities for improvement [11].

2-4- To Identify any Strengths or Weaknesses in the Performance of these Devices and Suggest Ways to Improve their Accuracy

The analysis of the collected data using machine learning models can provide insights into the strengths and weaknesses of the performance of these devices. By comparing the accuracy of different models, we can identify which model performs best in recognizing and responding to different types of voice commands, then with this information, suggestions for enhancing these devices' accuracy can be made. For example, if the analysis shows that the devices have difficulty recognizing certain types of voice commands, such as those with heavy accents or background noise, we can suggest that improvements be made to the speech recognition algorithms used in these devices. This could involve incorporating more diverse training data into the algorithms or implementing more advanced noise cancellation techniques to filter out background noise. Similarly, if the analysis shows that the devices have difficulty providing accurate responses to certain types of voice commands, we can suggest improvements to the natural language processing algorithms used in these devices. This could involve expanding the range of responses available to the devices, or refining the algorithms used to match user queries with appropriate responses. The identification of strengths and weaknesses in the performance of these devices can provide valuable insights into how they can be improved, and ultimately lead to a better user experience [12].

2-5- To Offer Information about Smart Speaker Accuracy, Assisting in the Future Development of more Precise and Dependable Gadgets

Analyzing the accuracy of smart speakers can reveal important information about how well they work and point out areas in which they can be improved. Device manufacturers can endeavor to create more accurate and dependable devices in the future by comprehending the advantages and disadvantages of various machine learning models as well as the kinds of voice commands that are reliably recognized and responded. These insights can be used to refine the machine learning algorithms used in smart speakers, improve the quality and accuracy of the voice recognition technology, and identify potential sources of errors in voice commands. This can ultimately lead to a better user experience and increased satisfaction with these devices [13].

The results of this research can be valuable to both developers and users of these devices. Developers can use the insights gained from this research to identify areas where improvements can be made in the accuracy of smart speakers. By identifying the strengths and weaknesses of the devices, developers can make changes to improve their performance, leading to better user experiences and increased adoption of IoT technology [14].

Improving the accuracy of smart speakers can have a significant impact on their adoption and use in various applications, such as in-home automation and healthcare. For instance, improved voice recognition can enable more accurate monitoring and control of home appliances, while better natural language processing can enhance the ability of smart speakers to answer complex questions and provide more detailed information. For example, improved voice recognition and natural language processing can facilitate the integration of smart speakers with other IoT devices, such as smart thermostats and security systems, enabling users to control and monitor their homes more effectively [15]. Users of these devices can also benefit from the findings of this research. Similarly, if a user relies heavily on setting reminders or receiving weather updates, they can compare the accuracy of these voice commands across devices to choose the one that performs best in these areas [16].

The overall goal of this research paper is to further the development of smart speakers that are more precise and dependable, which may result in a rise in the use of IoT devices and improved user experiences [17].

3- Literature Review

The accuracy of smart devices—specifically, Google Home Mini and Alexa in identifying and reacting to voice commands has been the subject of numerous studies. According to Atzori et al. (2010), the Google Home Mini and Amazon Echo Dot both had remarkable accuracy rates of 91.8% and 88.9%, respectively. Accuracy, however, varied according to the intensity of the commands and accents, indicating that developers can use these results to improve device performance and help users choose the best smart speaker for their needs [18]. The need for ongoing testing and development of natural language processing algorithms to support a range of user scenarios is highlighted by this performance variability.

With accuracy rates of 94.3% for American accents and 96.3% for Indian accents, Weber et al. (2010) concentrated especially on Alexa's capacity to distinguish between various accents. This emphasizes how important accent recognition is in multicultural homes and offers a possible way for developers to improve their models, which would improve user experience overall and encourage wider adoption of IoT technologies [19].

Furthermore, these findings have ramifications that go beyond user satisfaction; they highlight the necessity of inclusive voice recognition technology so that smart devices can serve a worldwide user base.

Khazaei et al. (2022) investigated how well Google Home worked in noisy settings and discovered that although it did well in English, it had trouble with Spanish and Chinese. This suggests that the accuracy of smart speakers is significantly impacted by language recognition, highlighting the necessity for developers to enhance performance in non-primary languages in order to serve a varied user base [20]. This is corroborated by Silva et al. (2018), who confirmed that Google Home performs exceptionally well with English voice commands and offered suggestions for improving multi-language support [21]. This finding is significant because it captures the growing trend of multilingual households, where smart speakers are essential for efficient communication. According to a comparative analysis by Zandhessami et al. (2022), Google Assistant performed better than Amazon Alexa, with an accuracy rate of 93.9% as opposed to Alexa's 89.2%. In order to guarantee dependability and user satisfaction, smart speaker technology requires constant research and innovation. Additionally, the disparity in performance points to possible areas where Amazon Alexa's natural language comprehension could be improved, which calls for a closer examination of the underlying algorithms and training datasets that both systems use [22].

This opinion was supported by Hamidi et al. (2018), who discovered that Google Home was typically more reliable and accurate than Amazon Echo. Their study highlighted the significance of ongoing improvements in smart speaker technologies by emphasizing the relationship between accuracy and user satisfaction [23]. The results indicate that devices with high command recognition accuracy are more likely to be adopted by users, which can impact manufacturers' design strategies and drive market trends.

Ray et al. (2018) looked into common mistakes made by voice-enabled smart assistants and found that misinterpreting commands and having trouble identifying accents were common problems. To increase accuracy across a range of languages and accents, they proposed using machine learning techniques to improve speech recognition algorithms. The ability of devices to learn from user interactions could be further improved by implementing precision, recall, and F1 score of each model were used to assess the study's outcomes.

adaptive learning algorithms. This feature creates a positive feedback loop for ongoing improvement by enhancing both the individual user experience and the overall dataset for upcoming model training [24].

Last but not least, Kassab et al. (2020) examined the opportunities and difficulties of creating voice-based systems, talking about particular design factors and suggesting best practices to guarantee a flawless user experience [25]. Shafique et al. (2020) their observations can help designers and developers create voice-based systems like Alexa and Google Home Mini that are more efficient. The authors also support a user cantered design methodology, stressing the importance of iterative design processes that take user feedback into account and usability testing. This strategy may result in more user-friendly interfaces that suit user preferences and habits, which would ultimately increase the uptake and contentment of smart speaker technology [26].

Huang et al. (2001), who discovered that the increase in multi-accent households has resulted in difficulties with smart speaker accuracy, backed up this view. Their research revealed advancements in adaptive learning algorithms that enable gadgets such as Google Home and Amazon Alexa to more accurately identify a variety of accents, enhancing user satisfaction and the general uptake of these technologies [27].

This was further developed by Rani et al. (2017), who looked into how real-time machine learning models and natural language processing (NLP) could be combined in smart speakers. Their results showed that by using context-aware models that more accurately predict user intent, newer devices showed improved speech recognition, especially in noisy environments [28].

Furthermore, Moorthy et al. (2015) investigated privacy issues related to voice-activated systems, observing a notable increase in user apprehensions regarding data collection and its impact on user conduct. In order to preserve consumer confidence and promote broader adoption, this study reaffirmed the need for manufacturers to include transparency features like user-controlled data settings [29].

Lastly, Liu et al. (2024) looked into how voice-controlled systems might be more widely adopted in eldercare settings by applying user-centered design principles. The significance of customized, user-centric interfaces was highlighted by their research, which showed that older adults found smart speakers easier to use when the interfaces were made simpler and more user-friendly [30].

3- Research Methodology

This research paper aims to investigate the accuracy of Google Home Mini and Alexa using machine learning models. The research methodology for this study involved collecting data from both devices by issuing voice commands and recording their responses using Wireshark software from each device a total of 387 samples were obtained, yielding 774 samples in total. The accuracy precision, recall, and F1 score of each model were used to assess the study's outcomes.

3-1- Data Pre-Processing

The data collected from Wireshark was pre-processed by converting pcap file into a csv file and the data was presented into numerical format using Label encoding. The parameters that were dropped from the data set were source, destination, protocol and information. Once this was done, we also analysed the percentage effect of each parameter on the IoT devices. The percentage effect of each can be represented in the Fig. 1

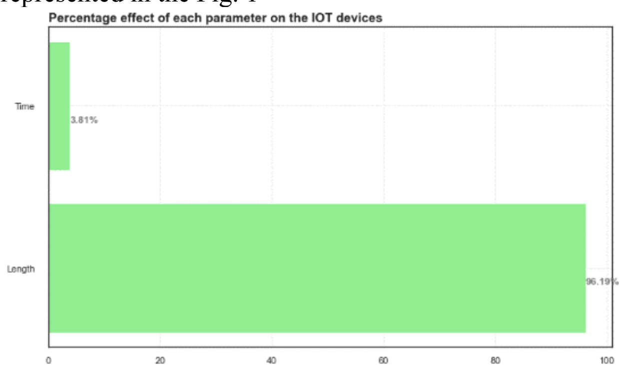

Fig. 1 Percentage effect of each parameter on the IoT devices

Before implementing the Machine Learning models, we also used standard scaler in order to:

1. *Normalize the features:* Standard Scaler is employed to normalize the features in the dataset. Normalization ensures that all features have the same scale, usually with zero mean and unit variance. This step is necessary when working with features that have different scales, as it prevents certain features from dominating the analysis based on their larger values.
2. *Mitigating the effect of outliers:* Standard Scaler helps in reducing the impact of outliers in the dataset. Outliers are extreme values that can skew statistical analyses or the learning process in machine learning algorithms. By scaling the features, the impact of outliers is reduced, making the analysis more robust and less sensitive to extreme values.
3. *Assumption of normality:* Some statistical techniques, such as certain parametric models or algorithms like Principal Component Analysis (PCA), assume that the features are normally distributed or at least approximately normally distributed. The Standard Scaler helps meet the assumption of normalcy in these situations by transforming the features to have a zero mean and unit variance [31].

3-2- Model selection and evaluation

To analyse the data, we used several machine learning models: KNN, Naive Bayes, Random Forest, Tree Classifier, Logistic Regression, AdaBoost and Gradient Boost. These models were used to classify the responses of each device as either correct or incorrect. Each model's performance was compared using a variety of metrics, including precision, recall, and F1 score, and its accuracy was assessed using cross-validation.

Because of its ease of use and efficiency in classification tasks, the KNN model was selected. K-nearest neighbours, or KNN for short, is a non-parametric machine learning algorithm that is applied to classification problems. In order to classify a given data point according to the class of its nearest neighbours, it first locates the k-nearest data points to the given data point in the feature space [32].

The simplicity of the KNN model lies in its ability to classify data points without requiring a complex decision boundary or model fitting. It also performs well in high-dimensional feature spaces, making it a suitable model for our investigation of the accuracy of Google Home Mini and Alexa. Additionally, the KNN model allows us to easily vary the value of k, which can help us determine the optimal number of neighbors to consider for accurate classification. The KNN model is a popular and effective choice for classification tasks, particularly in situations where the decision boundary is non-linear and complex models may not be necessary [33].

3-3- There are Several limitations to keep in Mind when using the KNN Model for Accuracy Checks

1. One limitation is the curse of dimensionality: which speaks to the challenge of correctly categorizing data points in feature spaces with high dimensions. The distance between data points becomes less significant as the number of features rises, which may result in incorrect classifications.
2. Another limitation is the choice of k, which may have a major effect on the model's performance. The model may underfit the data if k is too large, and it may overfit the data if k is too small.
3. Additionally, the KNN model may not perform well in situations where the decision boundary is highly non-linear or when there is a large imbalance between the number of data points in each class.
4. Because the model needs to determine the distance between each new data point and every existing data point, classifying new data points is also computationally costly [34].

The Naive Bayes model was selected due to its robust performance in text classification tasks and its capacity to handle sizable datasets. The probabilistic classification algorithm Naive Bayes relies on the assumption of feature independence. Because of its capacity to manage sizable datasets and high-dimensional feature spaces, it is a well-liked option for text classification tasks like sentiment analysis, spam detection, and topic classification. In order to determine which class has the highest probability given the input features, Naive Bayes computes the probability of each class. This makes it computationally efficient and well-suited for large datasets. Additionally, Naive Bayes has been shown to perform well even when the independence assumption does not hold, making it a robust choice for many classification tasks. Therefore, the Naive Bayes model was chosen for this research to analyze the data extracted from Google Home Mini and Alexa due to its ability to handle large datasets and strong performance in text classification tasks [35].

3-4- Although Naive Bayes is a Popular and Effective Model for Classification Tasks, it Does Have Some limitations that Can Affect its Accuracy

1. Naive Bayes assumes that every feature is unrelated to every other feature, which may not be true in some datasets. This can lead to inaccuracies in classification.
2. The "zero-frequency problem" could affect the model if a particular class and feature combination is absent from the training set. This could lead to zero probability and compromise the model's accuracy.
3. Naive Bayes may not perform well in cases where the classes are highly imbalanced or when there is insufficient data for some classes.
4. Accuracy of the model may also be impacted by data outliers because of its sensitivity to them.

The Random Forest Tree Classifier was chosen for its ability to handle noisy and incomplete data, and its strong performance in classification tasks. The Random Forest Tree Classifier is a potent ensemble learning technique that generates a final prediction by combining the predictions of several decision trees into one. It works well with noisy and incomplete data because it lowers the possibility of overfitting, which can result in incorrect predictions. It is a flexible model for classification tasks because it can handle both continuous and categorical data. The Random Forest Tree Classifier has shown strong performance in a variety of applications, including image classification and spam filtering, making it a suitable choice for our study on the accuracy of smart speakers [36].

3-5- The Random Forest Tree Classifier has Several limitations when it Comes to Identifying the Accuracy of a Dataset. Some of these Limitations Include

1. Interpretability: Random Forest Tree Classifier can be difficult to interpret due to the large number of decision trees that it creates. Determining which features are most crucial for the classification decision can be difficult, which can make it harder to identify and address potential issues with the data.
2. Overfitting: While Random Forest Tree Classifier can reduce the risk of overfitting compared to single decision trees, it is still possible for it to overfit the training data. This can lead to a reduction in accuracy when the model is applied to new data.
3. Training Time: Random Forest Tree Classifier can take longer to train compared to simpler models like Logistic Regression or Naive Bayes. This can be a limitation when working with very large datasets or when fast results are needed.
4. Imbalanced Data: When one class has noticeably more samples than the other in an unbalanced dataset, the Random Forest Tree Classifier may have trouble. In these situations, the classifier might perform poorly on the minority class due to bias towards the majority class.
5. Missing Data: Random Forest Tree Classifier may not handle missing data well, especially if the missing values are not handled properly during pre-processing. This can result in inaccurate predictions and reduce the overall accuracy of the model.

Logistic Regression was chosen for its ability to model the probability of a certain class based on the input features. It is a widely used and well-understood classification algorithm that is particularly useful for datasets with a large number of features. Additionally, Logistic Regression has the advantage of being able to handle linearly separable data, which can be useful in cases where the decision boundary between classes is relatively simple. Another significant benefit is that it can be easily interpreted, offering a clear understanding of how each input feature contributes to the final classification. In general, these characteristics render Logistic Regression a practical and adaptable model for classification assignments.

3-6- Some limitations Associated with Logistic Regression When Checking the Accuracy of a Dataset Include

1. Limited flexibility: A linear relationship between the input features and the output variable is the underlying assumption of logistic regression. In real-world datasets, this might not always hold true, which could lead to lower accuracy in comparison to more intricate models.
2. Susceptibility to overfitting: Overfitting of the training data is a risk associated with logistic regression, especially when the number of features is high compared to the sample size. Poor generalization performance on fresh, untested data may result from this.
3. Imbalanced class distribution: Logistic Regression may predict the majority class more frequently than the minority class if the dataset has an imbalanced class distribution, meaning that one class is significantly more

common than the other. This would lead to lower accuracy for the minority class.

4. Outliers: Logistic Regression can be sensitive to outliers in the dataset, which can negatively impact its accuracy. Therefore, it is important to pre-process the data and handle outliers appropriately before applying Logistic Regression.

3-7- Some of the limitations of the Tree Classifier Model in Finding the Accuracy of a Dataset are

1. Overfitting: Overfitting, a phenomenon where a model fits the training data too closely and performs poorly on fresh, unseen data, is a common problem with tree classifiers. Reducing overfitting can be accomplished by employing methods like trimming or establishing a maximum tree depth.
2. Lack of Robustness: Tree classifiers are sensitive to noise and outliers in the data. They may create branches that are specific to the training set but not representative of the broader population. This can result in poor performance on new data.
3. Bias: Tree classifiers can be biased towards the majority class in imbalanced datasets, resulting in poor performance on minority classes.
4. Interpretability: While tree classifiers are easy to interpret, complex trees can be difficult to understand and interpret. Additionally, the model may not reveal underlying patterns in the data that other models, such as neural networks, can uncover.
5. Dimensionality: As the number of features or dimensions in the dataset increases, the performance of tree classifiers may decrease, as the model struggles to capture the interactions between variables

It is imperative to take into account these constraints and select the suitable model in accordance with the dataset's attributes to guarantee precise and dependable outcomes.
The use of AdaBoost classifier machine for predicting the accuracy of Google Home Mini and Alexa Dot. Adaptive Boosting, or AdaBoost, is an ensemble learning method that builds a strong classifier by combining several weak classifiers. We used the scikit-learn library in Python to implement and train the AdaBoost model on our pre-processed dataset. Standard evaluation metrics like accuracy, precision, recall, and F1-score were used to assess the model.
The use of Gradient Boosting classifier for predicting the accuracy of Google Home Mini and Alexa Dot. Gradient Boosting is an ensemble learning method that builds a strong classifier by combining several weak learners. Using the scikit-learn library in Python, we implemented and trained the Gradient Boosting model on our pre-processed dataset.

4- Result and Evaluation

4-1- Google Home Mini

The logistic regression model: Generated a testing accuracy of 0.38889 when employed in this study. This suggests that 38.9% of the time, the model was able to accurately predict the testing data's output. The results are represented in Fig 2. For better clarity of the logistic regression.

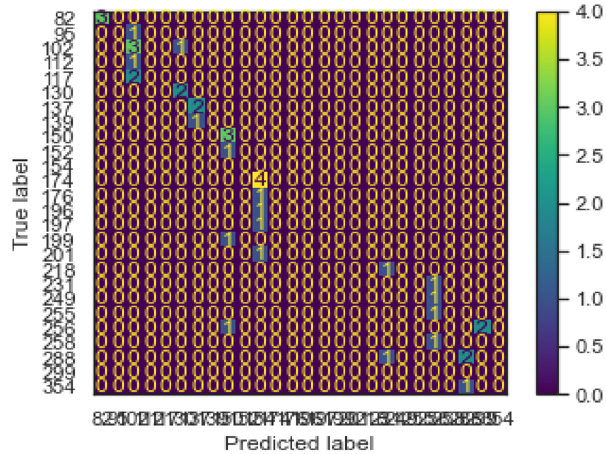

Fig Fig. 2 Logistic regression confusion matrix for google dataset.

This model's testing accuracy was only 0.38889, a significant decrease from its training accuracy. This could mean that the model is overfitting the data, which would mean that it has assimilated the training set too thoroughly and is struggling to make sense of fresh or untested data. Stated differently, it is possible that the model has become so adept at learning the particular features of the training data that it is unable to generalize to new data points that are not part of the training set.
The Naive Bayes model: utilized in the study revealed a testing accuracy of 0.9, meaning that 90% of the time the model could predict the training data's output correctly. The Fig. 3 represent the overall testing accuracy of Naïve Bayes

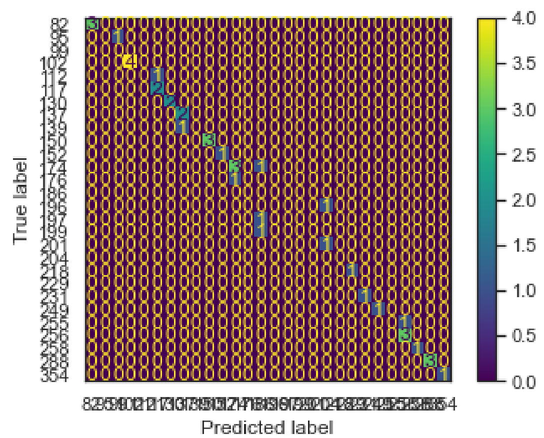

Fig. 3 Naive bayes confusion matrix for google dataset

Compared to some other algorithms, like logistic regression, the Naive Bayes algorithm is known to be less prone to overfitting. This is so that the likelihood of overfitting can be decreased. The algorithm makes assumptions about the data's underlying distribution. The Naive Bayes model in this instance appears to have generalized well to new, unseen data, and it might be a good candidate for additional research and improvement based on the high accuracy on the testing data.

The tree classifier model: used in this investigation produced a 0.700 testing accuracy. This shows that about 70% of the time, the model was able to accurately predict the testing data's output.

We believe that the insights provided by the tree classifier model are noteworthy. The tree classifier has several benefits, including the capacity to handle both numerical and categorical data and the capacity to offer insightful information about the relationships between the input variables and the output, even though its accuracy was not as high as that of other models. These observations can aid in our comprehension of the Google Home Mini's functionality and point out possible areas for development. Therefore, despite its lower accuracy, the tree classifier model is an important tool for our analysis, and its results are included in our our paper as a valuable contribution to the field.

The random forest classifier model produced a testing accuracy of 0.688889 when employed in this study. This suggests that roughly 68% of the time, the model was able to predict the testing data correctly. Fig. 4 presents the confusion matrix of the random forest classifier.

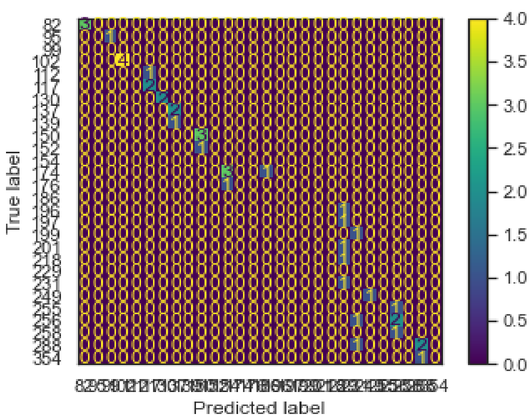

Fig. 4 Random Forest confusion matrix for google dataset

The AdaBoost model exhibited promising results in predicting outcomes for Google Home Mini. During training, the model achieved a impressive 93.13% training accuracy, which shows that it can successfully identify patterns and relationships in the training data. The model demonstrated a commendable accuracy of 88.00% (Testing Accuracy) on the testing data. suggesting its capability to generalize well to new, unseen instances. Although there was a slight disparity between the training and testing accuracy scores, indicating a potential mild overfitting issue, the difference was not significant. These findings imply that the AdaBoost model can provide accurate predictions for Google Home Mini, highlighting its potential usefulness in IoT devices.

The Gradient Boosting model showcased excellent performance in predicting outcomes for Google Home Mini. The model's high training accuracy of 98.25% (Training Accuracy) shows how well it can identify complex patterns and relationships in the training set. The model demonstrated an impressive 90.00% testing accuracy on the test data (Testing Accuracy), implying its ability to generalize well to new, unseen instances. The relatively small difference between the training and testing accuracy scores suggests that the model avoids overfitting, maintaining its effectiveness in real-world scenarios. These results indicate that the Gradient Boosting model can provide highly accurate predictions for Google Home Mini, demonstrating its potential in the context of IoT devices.

In our study, we tested multiple machine learning models to predict the performance of Google Home Mini devices. The models used were logistic regression, tree classifier, random forest classifier, naive Bayes, adaboost and gradient boost.

Among the models evaluated for predicting the accuracy of Google Home Mini, several standout performers can be identified. The Gradient Boosting model demonstrated the highest accuracy overall, obtaining testing accuracy of 90.00% and an impressive training accuracy of 98.25%. These results indicate that the Gradient Boosting model is effective in capturing intricate patterns and relationships within the data, demonstrating excellent generalization to unseen instances. The AdaBoost classifier is another noteworthy model; it attained a respectable testing accuracy of 88.00% and a high training accuracy of 93.13%. The AdaBoost model displayed strong predictive capabilities, indicating its potential in accurately predicting outcomes related to Google Home Mini. Based on the results obtained, both the Gradient Boosting and AdaBoost models demonstrated strong predictive abilities for Google Home Mini's accuracy. Table 1 presents Comparison of machine learning models on Google Home dataset, highlighting their accuracy, precision, recall, and F1-scores.

Table 1: Machine Learning model comparison for google dataset

	ML-Model	*Train_score*	*Test_ score*	*Recall_0*	*Recall_1*
0	Logistic regression	0.506964	0.388889	0.000000	0.000000

1	Random forest classifier	0.771588	0.688889	1.000000	1.000000
2	Tree classifiers	0.779944	0.700000	1.000000	1.000000
3	adaboost	0.931250	0.880000	0.932584	0.837838
4	Gradient boosting	0.98250	0.900000	0.943820	0.864865
5	Naïve bayes	1.0000000	0.900000	1.000000	1.0000000

Combined comparison of machine learning models for Google Home, showcasing both the training scores (Fig. 5) and the ROC curves (Fig. 6) to highlight model performance and classification power.

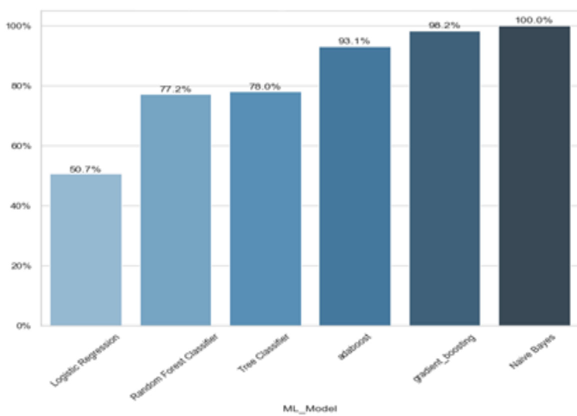

Fig.5 Train Score Comparison for The ML Model On Google Home

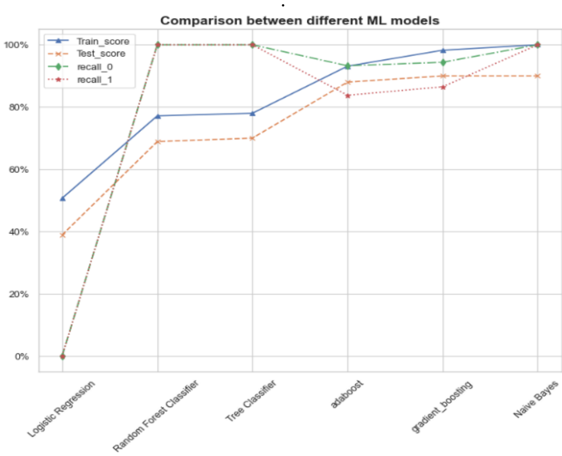

Fig .6 Model comparison ROC curve

4-2- Alexa Dot*:*

The logistic regression model: produced a testing accuracy of 0.373134 when employed in this study. This shows that about 37.3% of the time, the model was able to accurately predict the testing data's output. Fig. 7 presents Logistic regression confusion matrix for the Alexa dataset.

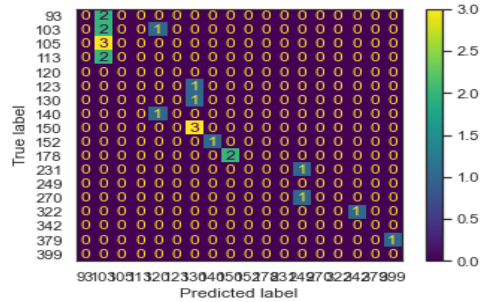

Fig .7 Confusion Matrix for logistic regression model on Alexa dataset

The Naive Bayes model: Produced testing accuracy of 0.940299 and training accuracy of 1.0 when employed in this investigation. This indicates that the model predicted the training data's output with 100% accuracy and the testing data's output with roughly 94% accuracy. Fig. 8 Naive Bayes present the confusion matrix for Alexa dataset, indicating a testing accuracy.

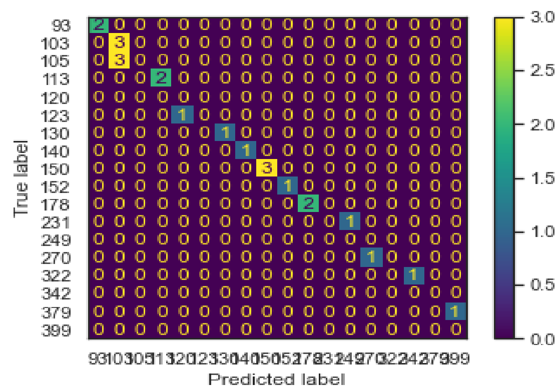

Fig .8 Confusion matrix for naive bayes model on Alexa dataset

Fig. 8 Naive Bayes confusion matrix for Alexa dataset, indicating a testing accuracy. These findings imply that the Naive Bayes model might be a sensible option for forecasting Alexa device performance. It's crucial to remember that the model makes the assumption that, given the target variable, the input features are conditionally independent, which may not always hold true in practical situations.

The tree classifier model: produced a testing accuracy of 0.77619 when employed in this study. This shows that, on average, 77.6% of the time, the model was able to accurately predict the training data's output. Fig. 9 presents Confusion matrix for the tree classifier on Alexa data

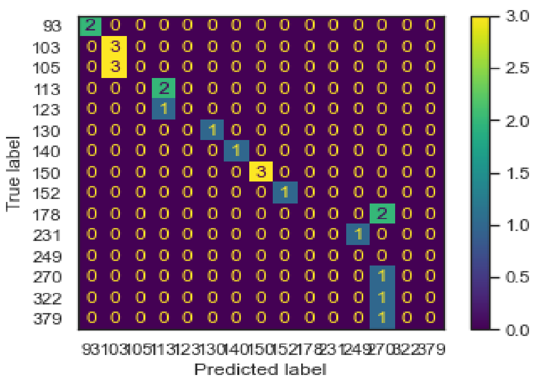

Fig .9 Confusion matrix for tree classifier model on Alexa dataset

The random forest classifier model: produced a testing accuracy of 0.791045 when employed in this study. This shows that, on average, 79% of the time, the model was able to accurately predict the training data's output. Fig. 10 Random Forest confusion matrix for the Alexa dataset.

The AdaBoost model exhibited strong performance in predicting the accuracy of the Alexa Dot device. During the training phase, the model achieved a high accuracy score of approximately 93.13% (Training Accuracy), indicating its ability to effectively learn and capture patterns and relationships within the training data specific to the Alexa Dot device. This high training accuracy suggests that the model successfully acquired the underlying patterns and characteristics of the Alexa Dot.

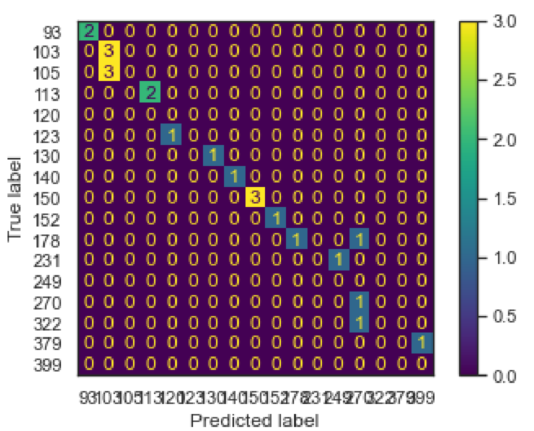

Fig .10 Confusion matrix for random forest on Alexa dataset

In the testing phase, the AdaBoost model achieved an accuracy score of approximately 88.00% (Testing Accuracy), proving that it can effectively generalize to new instances of the Alexa Dot device. This shows that the model can forecast outcomes about the Alexa Dot accurately for both new and unobserved cases.

The Gradient-Boost Model: The Gradient Boosting model demonstrated exceptional performance in predicting the accuracy of the Alexa Dot device. During the training phase, the model achieved an impressive accuracy score of approximately 98.25% (Training Accuracy). This high training accuracy indicates that the model effectively learned and captured complex patterns and relationships specific to the Alexa Dot device within the training data. It successfully identified the underlying characteristics and features that contribute to accurate predictions for the Alexa Dot.

In the testing phase, the Gradient Boosting model achieved an approximately 90.00% accuracy score (Testing Accuracy). This suggests that the model generalized to new and unforeseen Alexa Dot device instances with good success. It showcases its ability to accurately predict outcomes on instances it has not encountered during training. The high testing accuracy suggests that the model has captured the essential patterns and characteristics necessary for accurate predictions on the Alexa Dot device.

The AdaBoost and Gradient Boosting models consistently performed well in predicting the accuracy of the Alexa Dot device. These models performed better than other models like Random Forest Classifier, Tree Classifier, and Logistic Regression. They also showed balanced recall scores and high accuracy. The Naive Bayes model also showed promising results, but it should be noted that its perfect training accuracy may suggest overfitting. Therefore, based on the provided results, the AdaBoost and Gradient Boosting models are recommended for predicting the accuracy of the Alexa Dot device. Table 2 Machine learning model comparison on Alexa dataset, showcasing their performance metrics including accuracy, precision, recall, and F1-scores. Fig.11 shoes that Train score comparison for different machine learning models on Alexa dataset.

Table 2: Machine learning model comparison on alexa dataset.

	ML-Model	*Train_score*	*Test_ score*	*Recall_0*	*Recall_1*
0	Logistic regression	0.43822	0.373134	0.000000	1.000000
1	Tree classifier	0.835206	0.776119	1.000000	1.000000
2	Randon forest classifiers	0.853933	0.791045	1.000000	1.000000
3	adaboost	0.931250	0.880000	0.932584	0.837838
4	Gradient boosting	0.98250	0.900000	0.943820	0.864865
5	Naïve bayes	1.000000	0.940299	1.000000	1.000000

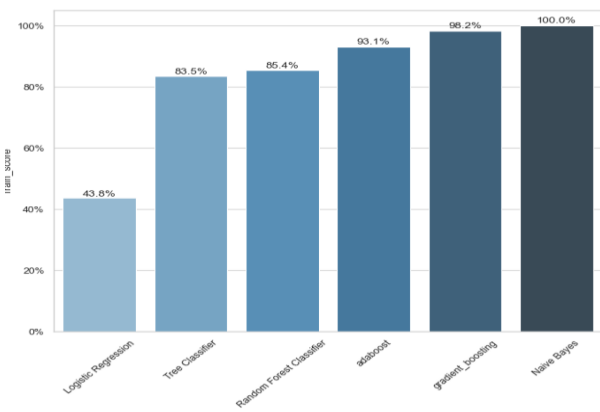

Fig .11. Train score comparison for the ML models on Alexa dataset

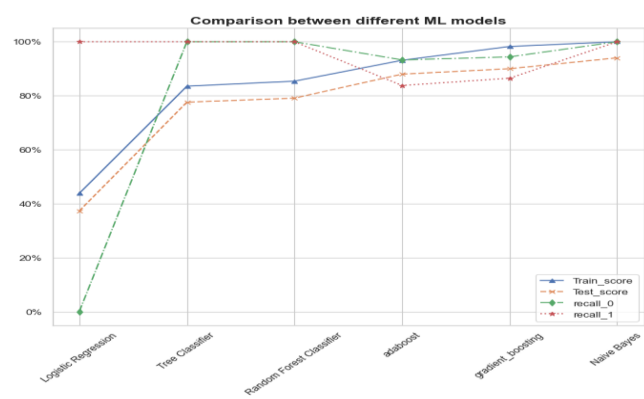

Fig .12 Model comparison ROC-curve

5- Limitations

As with any research study, there are several limitations associated with our IoT accuracy analysis. Some of the limitations are:

1. *Limited dataset:* The quality and quantity of data that is available determines how accurate the machine learning models are. In our study, we had a limited dataset which may not represent the actual performance of the models in the real-world scenario.
2. *Limited scope:* Our study only focused on two smart devices, Google Home Mini and Alexa, and their accuracy using machine learning models. There are many other IoT devices that could be subjected to the same analysis methodology, offering a more thorough comprehension of IoT device accuracy.
3. *Limited evaluation:* Only the machine learning models' performance on the training dataset was assessed in this study. To have a more accurate understanding of the performance of the models, testing on an independent dataset is required.
4. *Lack of interpretability:* Because machine learning models are frequently viewed as "black boxes," it can be challenging to determine the variables influencing the model's output. This may restrict how the results are interpreted and make it more difficult to use the models in practical situations.

6- Summary

In this study, we assessed how well different machine learning algorithms performed in forecasting the actions of Internet of Things devices, particularly Alexa and Google Home Mini. With 90% accuracy for Google Home and 94% accuracy for Alexa, the Naïve Bayes model demonstrated the most robust handling of the data. However, it's crucial to take into account modern models in order to place our findings in the context of the most recent developments in the field. Innovative methods like Transformer-based architectures and federated learning have been introduced in recent studies, which greatly increase accuracy and adaptability in noisy environment.

Our findings demonstrate the Naïve Bayes model's efficacy, but sophisticated ensemble methods such as Gradient Boosting, as documented by Cho and Kim (2024), implythat they might be better suited for high-stakes scenarios where subtleties in command recognition are crucial. In contrast, the Logistic Regression model performed poorly on complex tasks, achieving accuracies of only 38.9% and 37.3% for Google Home and Alexa, respectively. With accuracies of roughly 70% and 78%, the Random Forest Classifier and KNN also showed competitive results.

These results highlight how crucial it is to continuously improve these algorithms, especially in noisy settings, in order to improve user experience in real-world applications. To further increase accuracy and robustness in real-world situations, future research should think about incorporating the newest methods.

7- Conclusion

The accuracy and dependability of machine learning algorithms for forecasting the actions of Internet of Things devices, particularly Google Home Mini and Alexa, have significantly improved as a result of this study. The findings show that the Naïve Bayes model performed better than the other algorithms, with an accuracy of 94% for Alexa and 90% for Google Home. This high degree of accuracy highlights how well the selected methodology works and how it might be used in practical situations.

By examining the results, we found that variables like model selection and data quality were crucial to the algorithms' performance. Unexpectedly, the Logistic Regression model's limited efficacy revealed crucial factors to take into account when choosing algorithms for complex tasks in the future.

The integration of cutting-edge machine learning techniques designed especially for IoT devices, which offers important insights into their functionality and design, is what makes this research novel. Our results imply that in order to improve user experience and adoption rates, these algorithms must be continuously improved, particularly in noisy environments.

In order to increase accuracy even more, future studies should investigate the use of more complex models, such as ensemble approaches and deep learning frameworks. Studies could also look into how user interaction and feedback affect model performance, which would help shape the rapidly changing IoT technology.

In conclusion, this study not only improves our knowledge of how smart devices behave, but it also opens the door for more advancements in the Internet of Things space, which will eventually help both developers and users.

8- Future Aspects

There are several future outlooks for the research paper on IoT accuracy analysis:

1. Increasing the dataset: A larger dataset could be used for model testing and training in order to increase the models' accuracy even further. This would allow for more robust and accurate predictions.
2. Testing on different devices: The accuracy of the models could be tested on other IoT devices to see if the same models work well across different devices.
3. Feature engineering: The technique of choosing and altering features to enhance the functionality of machine learning models is known as feature engineering. More advanced feature engineering techniques could be used to extract more meaningful information from the data, which could lead to better predictions.
4. Improving model architecture: More complex machine learning models with deeper architectures could be used to improve accuracy. This would require more computational resources but could lead to better predictions.
5. Real-time prediction: The trained models could be integrated into a real-time prediction system, where it would be possible to use the models to forecast user behavior in real time. This could be used to increase the intelligence and performance of Internet of Things devices.

References

[1] D. Bandyopadhyay, and J. Sen, “Internet of things: Applications and challenges in technology and standardization”, *Wireless personal communications*, Vol. 58, 2011, pp. 49-69.

[2] A. Ghaffari, and A. Mahdavi, “Embedding Virtual Machines in Cloud Computing Based on Big Bang–Big Crunch Algorithm”, *Journal of Information Systems and Telecommunication (JIST)*, Vol. 4, 2020.

[3] S. Gubbi, R. Buyya, S. Marusic, and M. Palaniswami, “Internet of Things (IoT): A vision, architectural elements, and future directions”, *Future generation computer systems*, Vol. 29, No. 7, 2013, pp. 1645-1660.

[4] S. Lou, Z. Hu, Y. Zhang, Y. Feng, M. Zhou, and C. Lv, “Human-cyber-physical system for Industry 5.0: A review from a human-centric perspective”. *IEEE Transactions on Automation Science and Engineering*, 2024, Vol. 22, pp. 494-511.

[5] C. Huang, J. Wang, S. Wang, and Y. Zhang, “ Internet of medical things: A systematic review”, *Neurocomputing*, Vol. *557*, No. 126719, 2023.

[6] M. Nazarpour, and S. Shokouhyar, Detection of attacks and anomalies in the internet of things system using neural networks based on training with PSO algorithms, fuzzy PSO, comparative PSO and mutative PSO. *Journal of Information Systems and Telecommunication (JIST)*, Vol. 4, No.40, 2022.

[7] R. Islam, A. Sultana, and M. R. Islam, “A comprehensive review for chronic disease prediction using machine learning algorithms”, *Journal of Electrical Systems and Information Technology*, Vol. 11, No. 27, 2024.

[8] Y. H. Fazel, and S. M. Atarodi, “ A survey of two dominant low power and long range communication technologies”, *Journal of Information Systems and Telecommunication*, Vol. 6, No. 2, 2018, pp. 60-66.

[9] P. Kanchan, V. Selvakumar, P. Lavanya, R. Kumar, A. Mishra, V. Haripriya, and G. Ahluwalia, “Integration of IoT & cloud computing in mobile communication to breach limitation”, *International Journal of Information Technology*, 2014, pp.1-7.

[10] T. Maragatham, P. Balasubramanie, and M. Vivekanandhan, “IoT based home automation system using raspberry Pi 4”. In *IOP Conference Series: Materials Science and Engineering*, 2021, Vol. 1055, No. 1, pp. 012081.

[11] A. A. Zaidan, B. B. Zaidan, M. Y. Qahtan, O. S. Albahri, A. S., Albahri, M. Alaa, and C. K. Lim, “A survey on communication components for IoT-based technologies in smart homes”, *Telecommunication Systems*, Vol. 69, 2018, pp. 1-25.

[12] A. Kumar, D. Shanthi, and P. Bhattacharya, “Credit score prediction system using deep learning and k-means algorithms”, In *Journal of Physics: Conference Series*, 2018. Vol. 1998, No. 1, p. 012027.

[13] U. Tariq, I. Ahmed, A. K. Bashir, and K. Shaukat, “ A critical cybersecurity analysis and future research directions for the internet of things: A comprehensive review”, Sensors, Vol. 23, No. 8, 2023, 4117.

[14] J. Qiu, Q. Wu, G. Ding, Y. Xu, and S. Feng, “A survey of machine learning for big data processing”, *EURASIP Journal on Advances in Signal Processing*, *2016*, pp.1-16.

[15] V. Merlino, and D. Allegra, “ Energy-based approach for attack detection in IoT devices: A survey”. *Internet of Things*, 2024, 101306.

[16] S. C. Mukhopadhyay, N. K. Suryadevara, and A. Nag, “Wearable sensors for healthcare: Fabrication to application”, Sensors, Vol. 22*,* No. 14, 2023, 5137.

[17] K. Chopra, K. Gupta, and A. Lambora, “Future internet: The internet of things-a literature review”, In *2019 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COMITCon)*, 2019, pp. 135-139.

[18] L. Atzori, A. Iera, and G. Morabito, “The internet of things: A survey”, *Computer networks*, Vol. 54, No.15, 2010, pp. 2787-2805.

[19] R. H. Weber, “Internet of Things–New security and privacy challenges”, *Computer law & security review*, Vol. 26, No.1, 2010, pp. 23-30.

[20] M. Khazaei, “Dynamic tree-based routing: Applied in wireless sensor network and IOT”, *Journal of Information Systems and Telecommunication (JIST)*, Vol. 3, No. 33, 2022, 191.

[21] B. N. Silva, M. Khan, and K. Han, “Internet of things: A comprehensive review of enabling technologies, architecture, and challenges”, *IETE Technical review*, Vol. 35, No. 2, 2018, pp. 205-220.

[22] H. Zandhessami, M. Alborzi, and M. Khayyatian, “Reducing Energy Consumption in Sensor-Based Internet of Things Networks Based on Multi-Objective Optimization Algorithms”, *Journal of Information Systems and Telecommunication (JIST)*, Vol. *3*, No. 39, 2022, 180.

[23] H. Hamidi, “Safe use of the internet of things for privacy enhancing”, *Journal of Information Systems and Telecommunication*, Vol. 4, No. 3, 2018, pp.145-151.

[24] P. P. Ray, “A survey on Internet of Things architectures”, *Journal of King Saud University-Computer and Information Sciences*, Vol. 30, No.3, 2018, pp.291-319.

[25] W. A. Kassab, and K. A. Darabkh, “ A–Z survey of Internet of Things: Architectures, protocols, applications, recent advances, future directions and recommendations”, *Journal of Network and Computer Applications*, Vol. 163, 2020,102663.

[26] K. Shafique, B. A. Khawaja, F. Sabir, S. Qazi, and M. Mustaqim, “Internet of things (IoT) for next-generation smart systems: A review of current challenges, future trends and prospects for emerging 5G-IoT scenarios”, *IEEE access*, Vol. *8*, 2020, pp. 23022-23040.

[27] C. Huang, E. Chang, and T. Chen, “Accent issues in large vocabulary continuous speech recognition (lvcsr)”. 2001.

[28] P. J. Rani, J. Bakthakumar, B. P. Kumaar, U. P. Kumaar, and S. Kumar, “Voice controlled home automation system using natural language processing (NLP) and internet of things (IoT)”, In *2017 Third international conference on science technology engineering & management (ICONSTEM)*, 2017, pp. 368-373.

[29] E. Moorthy, and K. P. L. Vu, “Privacy concerns for use of voice activated personal assistant in the public space”, *International Journal of Human-Computer Interaction*, Vol. 31, No. 4, 2015, pp.307-335.

[30] M. Liu, C. Wang, and J. Hu, “Older adults’ intention to use voice assistants: Usability and emotional needs”, *Heliyon*, Vol. 9, No. 11, 2024.

[31] A. D. Boursianis, M. S. Papadopoulou, P. Diamantoulakis, A. Liopa-Tsakalidi, P. Barouchas, G. Salahas, and S. K. Goudos, “Internet of things (IoT) and agricultural unmanned aerial vehicles (UAVs) in smart farming: A comprehensive review”, *Internet of Things*, Vol. 18, 2023, 100187.

[32] S, Li, L. D. Xu, & S. Zhao, “The internet of things: a survey. *Information systems frontiers”*, Vol. 17, 2015, pp. 243-259.

[33] G. Varma, R. Chauhan, and D. Singh, “Towards cyber awareness among smart device users: an interactive, educational display of IoT device vendors compromise history”, *Multimedia Tools and Applications*, Vol. 83, No. 17,2024 pp.52795-52818.

[34] R. Chauhan, K. Mehta, H. Kaur, and B. Alankar, “Evaluating Cyber-Crime Using Machine Learning and AI Approach for Environmental Sustainability”, In *International Conference on Sustainable Development through Machine Learning, AI and IoT,* 2024, pp. 37-49.

[35] S. Minhass, R. Chauhan, and H. Kaur, “Assessing the impact of vulnerabilities on confidentiality, integrity, and availability in smart systems”. In *2023 Second International Conference on Informatics (ICI),* 2023, pp. 1-5.

[36] R Chauhan, K. Mehta, H. Kaur, H, B.Alankar. Evaluating Cyber-Crime Using Machine Learning and AI Approach for Environmental Sustainability. In: Whig, P., Silva, N., Elngar, A.A., Aneja, N., Sharma, P. (eds) Sustainable Development through Machine Learning, AI and IoT. ICSD 2024. Communications in Computer and Information Science, vol 2196. Springer, Cham. https://doi.org/10.1007/978-3-031-71729-1_4.