information

Article

Customer Shopping Behavior Analysis Using RFID and Machine Learning Models

Ganjar Alfian [1], Muhammad Qois Huzyan Octava [1], Farhan Mufti Hilmy [1], Rachma Aurya Nurhaliza [1], Yuris Mulya Saputra [1], Divi Galih Prasetyo Putri [1], Firma Syahrian [1], Norma Latif Fitriyani [2], Fransiskus Tatas Dwi Atmaji [3], Umar Farooq [4], Dat Tien Nguyen [5] and Muhammad Syafrudin [6,*]

[1] Department of Electrical Engineering and Informatics, Vocational College, Universitas Gadjah Mada, Yogyakarta 55281, Indonesia
[2] Department of Data Science, Sejong University, Seoul 05006, Republic of Korea
[3] Industrial and System Engineering School, Telkom University, Bandung 40257, Indonesia
[4] Faculty of Business and Law, Coventry University, Coventry CV1 5FB, UK
[5] Faculty of Electrical and Electronic Engineering, Phenikaa University, Yen Nghia, Ha Dong, Hanoi 12116, Vietnam
[6] Department of Artificial Intelligence, Sejong University, Seoul 05006, Republic of Korea
* Correspondence: udin@sejong.ac.kr

Abstract: Analyzing customer shopping habits in physical stores is crucial for enhancing the retailer–customer relationship and increasing business revenue. However, it can be challenging to gather data on customer browsing activities in physical stores as compared to online stores. This study suggests using RFID technology on store shelves and machine learning models to analyze customer browsing activity in retail stores. The study uses RFID tags to track product movement and collects data on customer behavior using receive signal strength (RSS) of the tags. The time-domain features were then extracted from RSS data and machine learning models were utilized to classify different customer shopping activities. We proposed integration of iForest Outlier Detection, ADASYN data balancing and Multilayer Perceptron (MLP). The results indicate that the proposed model performed better than other supervised learning models, with improvements of up to 97.778% in accuracy, 98.008% in precision, 98.333% in specificity, 98.333% in recall, and 97.750% in the f1-score. Finally, we showcased the integration of this trained model into a web-based application. This result can assist managers in understanding customer preferences and aid in product placement, promotions, and customer recommendations.

Keywords: shopping behavior; RFID; RSS; machine learning; outlier detection; data balancing

check for updates

Citation: Alfian, G.; Octava, M.Q.H.; Hilmy, F.M.; Nurhaliza, R.A.; Saputra, Y.M.; Putri, D.G.P.; Syahrian, F.; Fitriyani, N.L.; Atmaji, F.T.D.; Farooq, U.; et al. Customer Shopping Behavior Analysis Using RFID and Machine Learning Models. *Information* **2023**, *14*, 551. https://doi.org/10.3390/info14100551

Academic Editors: Christos Michalakelis, Mara Nikolaidou and Evangelia Filiopoulou

Received: 7 September 2023
Revised: 3 October 2023
Accepted: 6 October 2023
Published: 8 October 2023

1. Introduction

Comprehending customer behavior is essential for enterprises as it offers valuable insights into customer preferences and decision making, allowing companies to customize their offerings and marketing strategies to enhance customer satisfaction and loyalty [1]. While understanding customer shopping patterns in online stores is relatively straightforward, it becomes challenging in physical retail settings, where monitoring shopper behavior prior to checkout is complex. Utilizing Radio-Frequency Identification (RFID) technology is one approach to gaining insights into customer behavior.

RFID is a technology that uses wireless communication to identify and track objects or people using RFID tags or labels. These RFID tags contain unique identification data and are affixed to items, allowing them to be scanned and recognized by RFID readers or antennas [2]. RFID is regarded as an integral component of the Internet of Things (IoT) and finds extensive application across diverse industries, including inventory management [3], access control [4], asset tracking [5], and supply chain optimization [6], providing real-time data acquisition and enhancing operational efficiency.

RFID technology has emerged as a solution and has been implemented in physical stores for purposes such as smart trolleys [7–9] and analyzing customer shopping paths [10,11]. Additionally, previous research has demonstrated that the phase readings [12–14] and received signal strength (RSS) of RFID [15–17], combined with machine-learning algorithms, can effectively track customer activity within the physical store, including product browsing. Therefore, employing machine learning models for identifying customer behavior inside the store is crucial to enhance the efficiency of customer behavior analysis.

A multilayer perceptron (MLP), part of machine learning models, has shown success in enhancing system performance, particularly in classification tasks [18–22]. Yet, machine learning algorithms frequently encounter challenges such as outliers and imbalanced datasets, which can reduce accuracy. Research has demonstrated that addressing these issues by applying the Isolation Forest (iForest) method to remove outliers [23–27] and utilizing Adaptive Synthetic Sampling (ADASYN) for balancing imbalanced data [28–35] can lead to improved predictive system performance.

However, no prior research has explored the integration of iForest-based outlier detection, ADASYN, and MLP classifiers to enhance the accuracy of customer activity detection. Thus, this study suggests a predictive model that combines iForest-based outlier detection, ADASYN, and MLP to forecast customer activity based on time-domain features derived from RSS data. Additionally, incorporating this proposed predictive model into a web-based system can provide managers with a comprehensive view of product popularity and customer engagement. This data-driven approach enables managers to identify browsing patterns, potentially assisting in store layout optimization and overall shopping experience enhancement. The key contributions of this study can be summarized as follows:

- We introduced an innovative approach that combines iForest outlier detection, ADASYN data balancing, and Multilayer Perceptron to categorize customer activity, a novel contribution to the field.
- The model's performance was enhanced through the removal of outliers and balancing of the training dataset.
- We conducted comprehensive comparative experiments, comparing our model to other prediction methods and prior research findings.
- We analyzed the impact of outlier detection and data balancing methods, both with and without iForest and ADASYN, on the model's accuracy.
- Finally, we showcased the practicality of our model by designing and implementing a web-based customer behavior analysis system.

The subsequent sections of this study are structured as follows: Section 2 outlines the examination of customer behavior utilizing RFID and various machine learning models, encompassing iForest, ADASYN, and MLP. Section 3 details our suggested prediction model for detecting customer activity. In Section 4, we present the outcomes of our experiments and the execution of our model. Finally, in Section 5, we present the conclusions, including any study limitations and prospective avenues for further research.

2. Literature Review

This section introduced the utilization of RFID for customer behavior analysis, alongside the presentation of machine learning techniques such as iForest outlier detection, ADASYN data balancing, and Multilayer Perceptron.

2.1. Customer Behavior Analysis Based on RFID

Past studies have demonstrated the versatility of RFID technology, as it can be employed for various purposes, including enhancing smart trolley systems, tracking customer activity, and analyzing customer shopping activities within physical stores. In the context of smart trolleys, RFID technology reduces checkout wait times and automatically generates bills, eliminating the necessity for customers to queue at cashier counters. Badi and Momin [7] proposed a RFID and sensor-based system to reduce the waiting time for check out in superstores. In their proposed system they also implemented machine learning algorithms such as Eclat association rule to identify frequent items purchased and use it for sales promotion sent directly to customers smart phones. In another similar study, Athauda et al. [8] designed a Smart Trolley framework to address the challenges faced by retailers such as efficient stock maintenance and shoplifting. The proposed system has used UHF circular polarized RFID readers, including antennas and hybrid couplers to efficiently trace customers picked items in the trolley by overcoming previous limitations of orientation, size, and shape of tagged shopping items in the trolley. This RFID system allows for real-time tracking and processing of shopping data, providing consumers with instant access to information via an interface. Finally, Pradhan et al. [9] introduced KONARK, an RFID-based smart shopping system centered around a customized shopping cart equipped with an RFID reader. KONARK aims to expedite item checkout and enable real-time purchase tracking, providing users with a more efficient shopping experience. Additionally, the system accurately detects user interest in specific items and offers valuable insights to shopping mall owners, demonstrating robust performance across various mobility speeds in a simulated shopping mall setting.

RFID technology facilitates the monitoring of customer movements within the store, enabling the generation and retention of shopping routes for later examination. Previous studies have shown that this can provide valuable insights into customer behavior. Nakahara and Yada [10] explores the potential of shopping path data, generated by tracking customer movements in stores, to uncover insights into customer behavior and its impact on purchasing decisions. Using LCMseq analysis on RFID real data from a Japanese supermarket, the study reveals distinctive in-store behavior patterns associated with high-value customers. In a related study, Shen et al. [11] discusses the growing use of RFID technology, particularly in retail, to track in-store customer behavior. It introduces a unified framework for RFID-based path analytics, combining in-store shopping paths and RFID-based purchase data to extract actionable navigation patterns.

While comprehending customer shopping patterns in online stores is relatively straightforward, it poses a challenge in physical retail outlets, where monitoring shopper behavior before the checkout process is complex. Phase measurement from RFID has been applied to detect customer activities, resulting in noteworthy findings. Zhou et al. [12], utilized cyber physical system of RFID tags, and readers to investigate the shopping behavior of customers to come up with effective marketing strategies for the physical shopping stores. The proposed system ShopMiner system used customer's interactions with products and leveraged backscattered time-series phase readings from RFID tags to identify popular, hot, and correlated items. Liu et al. [13] presented TagBooth, an innovative system utilizing Commercial Off-The-Shelf (COTS) RFID devices to detect product motion and uncover customer behaviors, offering valuable insights beyond typical transaction data. They employed phase measurements as input data to distinguish between different forms of customer browsing. The system demonstrates robust performance in both laboratory and real retail store environments. Shangguan et al. [14] presented ShopMiner, a framework utilizing passive RFID tags' backscatter signals to detect customer activities such as browsing, examining, and trying on items. By leveraging unique spatial–temporal correlations in time-series phase readings, ShopMiner achieves high accuracy and efficiency in identifying comprehensive shopping behaviors, as demonstrated in empirical studies conducted in typical indoor environments.

In order to understand in-store customer behavior, including activities such product browsing, researchers have employed machine learning prediction models in conjunction with received signal strength (RSS) data, leading to significant discoveries. To tackle the architectural design and privacy challenge for automated checkout systems, Hauser et al. [15] explored the implementation of an automated checkout system in fashion retail stores, aiming to improve customer experiences and reduce operational costs. The study aims to detect products as they pass through the RFID gate based on machine learning methods, i.e., logistic regression, artificial neural networks, support vector machine, and gradient tree boosting. The system employed RFID tag RSS, and its successful application and assessment in a real-world context underscore its capability in effectively managing shopping baskets. Choi et al. [16] utilized RFID tag data to classify customer browsing patterns within specific zones. RFID devices capture customer behavior and preferences to inform business decisions and tailor marketing efforts, while intelligent fuzzy screening algorithms assist in matching apparel items according to customer preferences, product design, and sales history. In summary, the system is anticipated to enhance the shopping experience through smart, personalized services, ultimately contributing to the success of the retail business. Finally, Alfian et al. [17] focuses on analyzing customer shopping behavior, particularly browsing activities, in retail stores using RFID-enabled shelves and machine learning models. RFID technology is implemented to monitor tagged product movements, and a dataset is generated from the receive signal strength (RSS) of tags for different customer behavior scenarios. The results demonstrate that the proposed MLP-based model significantly outperforms other models in terms of accuracy, precision, recall, and f-score, offering valuable insights for product placement, promotions, and personalized recommendations to customers.

A limitation observed in earlier research lies in the challenge of understanding customer shopping patterns in physical retail environments, especially in monitoring shopper behavior prior to the checkout stage, which can be intricate. While some studies have utilized machine learning models with received signal strength (RSS) data, there is room for improvement in model accuracy. Enhancing the model's precision is achievable by integrating outlier detection methods and implementing data balancing techniques.

2.2. Isolation Forest Outlier Detection

Isolation Forest (iForest) is an outlier detection algorithm that works by isolating anomalies in a dataset. It does so by creating a binary tree structure that separates normal data points from outliers, making it efficient and effective for identifying data points that deviate significantly from the majority [36]. Numerous research investigations have shown that incorporating iForest can significantly enhance classification accuracy across various domains.

Within the framework of Industry 4.0, the utilization of iForest for outlier detection has found application in IoT scenarios, including Wireless Sensor Networks (WSN) and RFID systems. Ribeiro et al. [23] addresses the challenge of detecting abnormal screw tightening processes, a critical industrial task with cost-intensive manual labeling requirements. The research focuses on unsupervised detection methods, specifically isolation forest (iForest) and deep learning autoencoder (AE), and conducts extensive experiments using various datasets. Both iForest and AE demonstrate excellent anomaly detection performance, with IForest offering competitive results. Chen et al. [24] introduced a system with the goal of boosting the security and dependability of IoT systems by efficiently detecting irregularities in sensor data acquired from WSNs. They introduced a method called BS-iForest (box plot-sampled iForest) tailored for wireless sensor networks, which is based on a modified version of the Isolation Forest algorithm. The results of this approach demonstrated enhanced performance and stability, presenting a hopeful solution to address the security challenges encountered by IoT systems. Alfian et al. [25] focuses on utilizing machine learning algorithms to detect the movement and direction of passive RFID tags within a supply chain. The dataset includes various conceivable tag motions and directions, simulating real warehouse scenarios. The proposed model, combining Isolation Forest (iForest) outlier detection, Synthetic Minority Over Sampling Technique (SMOTE), and Random Forest (RF), achieves a high accuracy of up to 94.251% in detecting RFID tag movement and direction, outperforming other machine learning models.

Prior studies have demonstrated the influence of the iForest technique on enhancing prediction model performance within the domains of real urban 3D reconstruction and electronics manufacturing. Li et al. [26] have presented a novel approach to extract and clean street tree trunk point cloud data acquired through Mobile Laser Scanning (MLS). This method is built upon an improved version of the Isolation Forest (iForest) algorithm and focuses on enhancing the accuracy of trunk extraction and noise reduction. In comparison to conventional denoising techniques such as the Statistical Outlier Removal (SOR) filter and the Radius filter, the proposed method stands out with an impressive 30% enhancement in denoising accuracy, especially for noise points located near tree trunks. Zeng et al. [27] focuses on the application of isolation forest (iForest) in quality monitoring for micro resistance spot welding (MRSW) used in electronics manufacturing, particularly for joining micro enameled wires. The iForest-based anomaly detection model effectively distinguishes incomplete fusion welds from normal ones and can enhance online quality monitoring in enameled wire welding processes in production.

2.3. ADASYN Data Balancing

ADASYN, or Adaptive Synthetic Sampling, is a data resampling technique used in machine learning for addressing class imbalance issues. It generates synthetic samples for the minority class with a higher focus on those instances that are more challenging to classify correctly, thereby helping improve the model's performance on imbalanced datasets [37]. Many research studies have demonstrated that the utilization of ADASYN can substantially improve the accuracy of classification in many areas.

ADASYN was implemented on the health dataset, leading to notable and meaningful outcomes. Javeed et al. [28] proposed a machine learning model for dementia prediction that addressed bias by incorporating synthetic sampling through the ADASYN technique. The model utilized the Feature Extraction Battery (FEB) for feature extraction and a fine-tuned Support Vector Machine (SVM) with a radial basis function (RBF) kernel to enhance accuracy. Following the data split, the ADASYN technique was implemented on the training data, ensuring that the proposed system (FEB-SVM) was trained using a balanced dataset. The novel FEB-SVM model resulted in a 6% improvement in dementia prediction performance compared to the traditional SVM. Askari et al. [29] explores the potential for precision medicine in treating chronic diseases such as diabetes by detecting and classifying acute psychological stress (APS) and physical activity (PA) using wristband-collected physiological data. It employs random convolutional kernel transformation to create feature maps from biosignals such as blood volume pulse, galvanic skin response, skin temperature, and accelerometer data. The combination of partial least squares–discriminant analysis (PLS-DA) for feature selection and adaptive synthetic sampling (ADASYN) for data balancing demonstrates the best overall performance, offering promising precision medicine prospects for diabetes treatment. Finally, Raihan et al. [30] explores the use of smartphone technology to detect the presence of COVID-19, given the challenges and risks posed by the pandemic. To address the class imbalance issue in the COVID-19 dataset, Adaptive Synthetic Sampling (ADASYN) is employed. The study utilizes feature selection methods to identify ten optimal features from the dataset and applies various classification algorithms, with eXtreme Gradient Boosting (XGBoost) achieving the highest accuracy at 97.91%.

Furthermore, numerous studies in the field of network intrusion detection have employed ADASYN and have demonstrated positive outcomes in enhancing model accuracy. Zakariah et al. [31] addresses the inadequacy of traditional security measures in IoT networks by proposing a Long Short-Term Memory (LSTM) model for intrusion detection. The study applies adaptive synthetic sampling (ADASYN) to augment minority-class samples, enhancing the model's ability to identify all classes accurately. Experimental results on a benchmark dataset demonstrate the model's improved accuracy, precision, recall, and F1 score, outperforming previous algorithms and showcasing its potential for bolstering IoT network security. Cao et al. [32] presents an intrusion detection model that combines a convolutional neural network and a gated recurrent unit, addressing issues related to low accuracy and class imbalance in existing intrusion detection models. To tackle the sample imbalance problem, a hybrid sampling algorithm involving Adaptive Synthetic Sampling (ADASYN) and Repeated Edited nearest neighbors (RENN) is applied. Experimental results on multiple datasets demonstrate significant improvements in classification accuracy, effectively mitigating the challenges of low accuracy and class imbalance in intrusion detection. Finally, Fu et al. [33] proposed a deep learning model for network intrusion detection (DLNID) which addressed data imbalance issues by applying adaptive synthetic sampling (ADASYN). The ADASYN used for sample expansion of minority class samples, aiming to create a more balanced dataset. This result was significantly better than the ones that was not processed using any data augmentation method or processed using SMOTE method.

Moreover, ADASYN has been utilized to address data imbalance issues, resulting in improved accuracy for predicting students' performance and detecting bridge defects. Thaher et al. [34] addresses the challenge of predicting students' performance by introducing an evolutionary-based model utilizing an enhanced form of the Whale Optimization Algorithm (EWOA) for wrapper feature selection. The study employs the Adaptive Synthetic (ADASYN) sampling technique to handle imbalanced data, enhancing the prediction quality. Li et al. [35] aimed to assist bridge managers in making informed maintenance decisions for steel bridge deck defects by developing an ensemble-learning-based prediction model. The research employed ADASYN to address data imbalance issues in the 2021 NBI database and built six ensemble learning models with optimized hyperparameters through grid search. XGBoost emerged as the optimal model, boasting high accuracy (0.9495), AUC (0.9026), and F1-Score (0.9740).

2.4. Multilayer Perceptron

MLPs are designed for supervised learning tasks and possess the ability to make predictions across a wide range of fields. The application of MLP in the healthcare domain resulted in a significantly elevated level of prediction accuracy. Pasanisi and Paiano [18] used MLP to predict Cardiovascular Disease (CVD) risk in patients. The result of K-cross validation with 10 iterations showed that 94.91% were correctly classified. It indicates that the proposed method is an efficient strategy in predicting the risk of chronic diseases such as CVD. Javed et al. [19] introduced a proposed system designed as a predictive tool for evaluating the risk of depression and anxiety during pregnancy, specifically termed as antenatal depression and antenatal anxiety. This system employs a Multi-Layer Perceptron-based Neural Network (MLP-NN) classifier, integrating feature selection using ReliefF. The findings reveal that the MLP-NN attained an area under the receiver operating characteristic curve of 88% for antenatal depression and 85% for antenatal anxiety, outperforming the support vector machine classifier, which achieved 80% for antenatal depression and 77% for antenatal anxiety.

MLP has been utilized in manufacturing and energy consumption contexts, yielding notable outcomes. Ke and Huang [20] tried to enhance the quality control procedure for injection-molded items by introducing a multilayer perceptron (MLP) neural network model integrated with quality indicators to predict the dimensions of the final products precisely and swiftly. The outcomes exhibited exceptional accuracy, surpassing 92%, in forecasting the width of the geometric attributes using the specified quality indicators. This system has the potential to bring substantial advantages to the injection molding sector by automating inspections and elevating inspection accuracy. Ferreira et al. [21] proposed MLP as an approach to estimate the impact of data center energy consumption on the environment. In this research, MLP is applied to the software for the energy flow model (EFM), which computes the energy requirements of the data center. Furthermore, based on the computation of energy demand, the proposed MLP model of this study is capable to forecast the amount of CO_2 emissions caused.

Ultimately, in the agricultural context, MLP can be effectively employed for achieving highly accurate predictions. Ahmed [22] aimed to tackle the important challenge of predicting crop yields, essential for informed agricultural decision making at various geographic levels. It utilizes a blend of the Multi-Layer Perceptron (MLP) model and the Spider Monkey Optimization (SMO) technique. By using the Crop Yield Prediction Dataset, it forecasts the maize yield in Saudi Arabia, considering factors such as temperature, rainfall, and past production. The proposed MLP-SMO model outperforms other approaches, demonstrating its potential in providing accurate crop yield predictions.

Prior studies have indicated that Multilayer Perceptron (MLP) models are applicable across various domains for predictive modeling. Nevertheless, these studies have highlighted that the MLP model's performance may be compromised when dealing with outliers and imbalanced dataset. It is anticipated that by eliminating outliers and achieving a balanced training set, the MLP model's classification accuracy will be enhanced.

3. Methodology

In this study, we aim to identify customer shopping behavior using machine learning models. The machine learning model depicted in Figure 1 utilizes RSS to predict customer shopping patterns. To prepare the dataset, inconsistent entries were removed through preprocessing, and missing values were substituted with the mean. The RFID readings dataset underwent outlier detection employing iForest, facilitating the removal of anomalies. Extracted from the RSS data were time-domain features, while the ADASYN technique was implemented to create new instances of the minority class. For prediction, the MLP algorithm was employed, and model efficacy was assessed by contrasting it with alternative machine learning models. The model's assessment employed stratified 10-fold cross-validation, a variant of k-fold cross-validation that maintains the original dataset's class distribution in each subset. Lastly, the trained model was integrated into a web-based application, enhancing accessibility for end users.

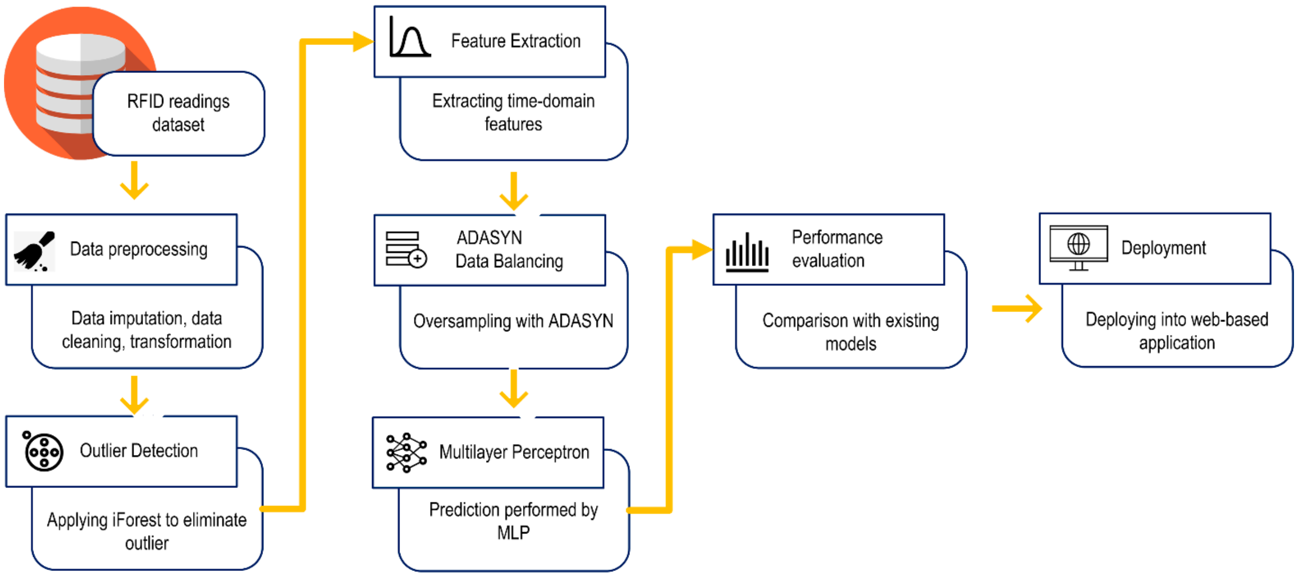


Figure 1. Proposed ML model to detect customer shopping activity.

3.1. Dataset

In our study, we investigated the shopping behavior of customers within a retail store, specifically examining their actions while browsing products. To initiate this study, we introduced an RFID-enabled shelf within the retail establishment. This shelf consists of a singular RFID reader and an antenna oriented directly towards the products, where passive tags were attached. Following this, we collected the receive signal strength (RSS) emitted by the tags across various customer behavior scenarios [17]. We then derived time-domain attributes from the RSS data of the tags. Subsequently, we applied a range of machine learning algorithms to differentiate between distinct customer behaviors concerning the tagged items. We conducted the experiments in a controlled laboratory setting, simulating a typical scenario within a retail store. Figure 2a outlines possible customer behaviors that may occur around products, such as a customer browsing the product or not exhibiting any interest (lack of attention to the product). Furthermore, Figure 2b presents an illustration of a tagged product being browsed by a customer, while Figure 2c portrays our developed data gathering program obtaining RSS data from RFID tags in real-time. In our experimental setup, customers took less than 15 s to browse each product, and during each session, we collected the RSS data from the tagged products for further analysis.

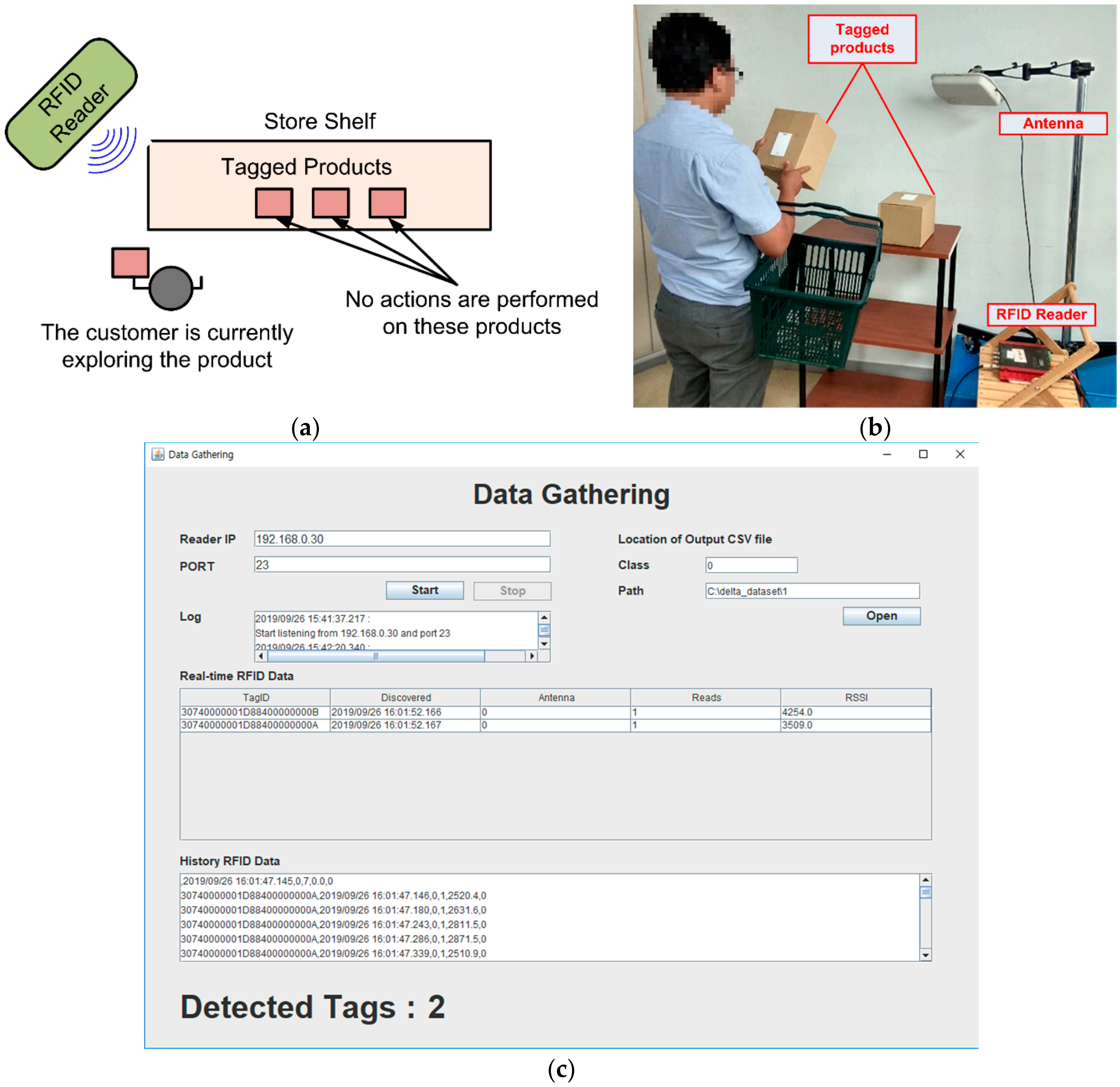


Figure 2. A potential situation illustrating (**a**) customer purchasing patterns and (**b**) an instance where a customer is exploring the product, and (**c**) data gathering program to record RFID tag RSS.

We examined two kinds of tag reads, "no action" and "browsing" tags. The "no action" tag signifies instances where the customer shows no interest in the product, whereas the "browsing" tag indicates that the customer is actively exploring the product. This study employed a single RFID reader, the ALR-9900+, along with linear antennas ALR-9610-AL having a Gain of 5.90 dbi [38]. Furthermore, all products were attached with UHF passive tags (Alien H3 model 9662). We improved additional experiment in our previous study [17], finally in total, 117 distinct data readings were gathered, with 70 readings categorized as "no action" and 47 readings as "browsing".

Figure 3 illustrates an instance of RSS readings depicting a typical shopping behavior of a customer within a retail store. In the scenario of "no action," the RSS of the tag remains relatively constant due to the fixed distance between the antenna and the tagged product, as seen in Figure 3a. Conversely, Figure 3b illustrates a situation where the customer is browsing the product. During this process, the RSS diminishes as the tag moves away from the antenna, reaching its lowest point when the tag is farthest from the antenna or closest to the customer. When the product is returned to its original position, the RSS value increases and stabilizes once more. In contrast to the "no behavior" tag, the "browsing" tag typically displays a greater variation in RSS. However, real-world scenarios can encompass various circumstances. As demonstrated in Figure 3c, the signal from the "no action" or stationary tag to the antenna can be obstructed by the movement of other products. This obstruction causes the RSS values to decrease as other products come in between the line of sight (LOS) of the tag and the antenna. Lastly, Figure 3d illustrates a scenario where a tagged product is initially browsed by a customer but is subsequently returned to a different location farther from the antenna. In this case, the RSS value decreases as the tag moves away from the antenna, resulting in a lower RSS reading when the tag is positioned farther from the antenna. These conditions contribute to the creation of a complex dataset representing various customer shopping behaviors. To effectively classify these behaviors, suitable attributes need to be extracted from the time series dataset and used as input for classifiers.

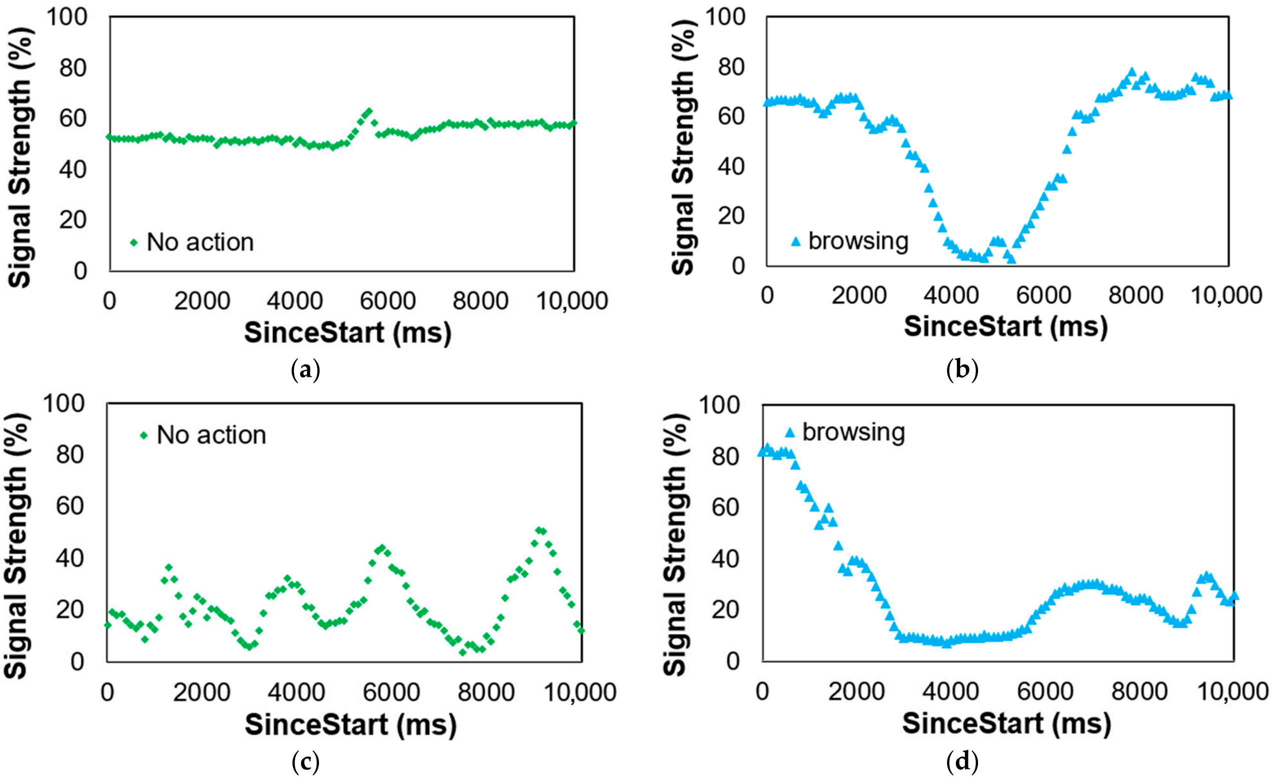


Figure 3. Instances of RSS readings: (**a**) the customer does not interact with the product, (**b**) the customer engages with the product by browsing, (**c**) the signal from an unmoved product is obstructed by the movement of another objects, and (**d**) the customer browses the product and then returns it to a different location.

3.2. Feature Extraction

RSS information relies on the proximity of the tag to the antenna, therefore, tags that are closer produce higher RSS values. The time-domain features of RSS crucial for distinguishing between tags that have moved and those that have not, as shown in earlier research [17,39,40]. In each collection session, we derive time-domain characteristics from the RSS data, including metrics such as Minimum, Maximum, Mean, Standard Deviation, Difference (indicating signal strength range), Kurtosis, Skewness, Entropy, and Count (total RFID tag read occurrences). As depicted in Figure 3, these characteristics would be beneficial for distinguishing between customer behaviors such as "no action" and "browsing". In addition, the choice of these specific time-domain features is justified by their proven effectiveness in capturing relevant information for our analysis. They have been shown in previous research to be well-suited for discerning between various tag movements and conditions. Furthermore, we acknowledge that the selection of these features is not arbitrary but based on their demonstrated value in similar studies [17,39,40].

The initial processing phase must be carried out to transform tag measurements into an input matrix X and an output vector Y. This enables traditional machine learning models to grasp patterns and foresee results. Ultimately, with n representing the count of distinct tag instances during each data collection session, along with a total of 10 time-domain features and m unique tag measurement data, the input matrix X can be formulated as a $[m \times 10]$ matrix.

$$X = \begin{bmatrix} \min(RSS_{1,1}, \ldots, RSS_{1,n}) & \max(RSS_{1,1}, \ldots, RSS_{1,n}) & \ldots & \mathrm{count}(RSS_{1,1}, \ldots, RSS_{1,n}) \\ \min(RSS_{2,1}, \ldots, RSS_{2,n}) & \max(RSS_{2,1}, \ldots, RSS_{2,n}) & \ldots & \mathrm{count}(RSS_{2,1}, \ldots, RSS_{2,n}) \\ \vdots & \vdots & \vdots & \vdots \\ \min(RSS_{m-1,1}, \ldots, RSS_{m-1,n}) & \max(RSS_{m-1,1}, \ldots, RSS_{m-1,n}) & \ldots & \mathrm{count}(RSS_{m-1,1}, \ldots, RSS_{m-1,n}) \\ \min(RSS_{m,1}, \ldots, RSS_{m,n}) & \max(RSS_{m,1}, \ldots, RSS_{m,n}) & \ldots & \mathrm{count}(RSS_{m,1}, \ldots, RSS_{m,n}) \end{bmatrix} \quad (1)$$

Each individual set of tag readings corresponds to customer behavior, categorized as either "no action" or "browsing," denoted as $Class_m$. Ultimately, the desired output Y was structured as a $[m \times 1]$ vector.

$$Y = \begin{bmatrix} Class_1 \\ Class_2 \\ \vdots \\ Class_{m-1} \\ Class_m \end{bmatrix} \quad (2)$$

3.3. iForest Outlier Detection

The iForest technique operates by forming a collection of isolation trees (*iTrees*) for each dataset, categorizing outliers as instances with notably brief average path lengths within the *iTrees* [36]. These *iTrees* are progressively established by partitioning the dataset until either all instances are isolated or a predefined tree height is attained. The pseudocode for iForest is outlined in Algorithm 1. Given input data D, *MaxSample* (representing the size of subsamples), and *NumTree* (indicating the quantity of trees to construct), the algorithm produces numerous *iTrees* and yields a Forest. Each *iTree* is created using sampled data from D.

Algorithm 1: Isolation forest.

Input		*MaxSample*, *NumTree*, D
Output		Collection of *iTrees*
1:	create Forest	
2:	establish a height restriction $h = \mathrm{ceiling}(\log_2(\ MaxSample))$	
3:	for i = 1 to *NumTree* do	
4:		$D' \longleftarrow$ sample(D, *MaxSample*)
5:		Forest $\longleftarrow$ Forest $\cup$ *iTree* (D', 0, h)
6:	end for	
7:	return Forest	

During the generation of *iTrees*, the subsequent step involves the random selection of a feature q from each *iTree*, along with the random selection of a value p within a given range. Subsequently, the data are divided into two segments: a left branch containing data points where $q < p$, and a right branch comprising data points where $q \geq p$. This procedure is repeated until only one data point remains in a branch or until the *iTree* reaches its maximum depth. This sequence is reiterated numerous times for each *iTree*, culminating in the creation of a Forest. The concluding phase encompasses identifying data points from each *iTree* that possess shorter path lengths, designating them as outliers. In this study, 18 outlier data points were detected and removed from the RFID readings dataset. Ultimately, the remaining dataset consisted of 99 instances: 64 instances labeled as "no action" and 35 instances labeled as "browsing". This refined dataset was then employed for subsequent analysis.

3.4. ADASYN

ADASYN, known as Adaptive Synthetic Sampling, is a technique employed in machine learning to address data imbalance, especially when there is a disproportionate distribution of instances across different classes in a training-set [37]. The core process of ADASYN involves several steps. First, it calculates the imbalance ratio, quantifying the disparity between the majority and minority classes. For each instance in the minority class, ADASYN identifies its k-nearest neighbors within the majority class, with 'k' determined according to the calculated imbalance ratio. Following this, ADASYN computes the density difference between each minority instance and its corresponding k-nearest neighbors. This density difference indicates the level of imbalance in the neighborhood and highlights regions where the minority class is underrepresented. By emphasizing instances in areas of higher density difference, ADASYN ensures that synthetic samples are strategically placed to improve the minority class's representation.

Subsequently, ADASYN generates synthetic samples for the minority instances, with the number of samples determined based on the density difference. Instances located in regions with more pronounced density differences receive more synthetic samples, fostering a balanced representation. This approach is particularly effective in addressing areas where the class imbalance is most severe. To achieve an optimal balance, ADASYN repeats the entire process iteratively for each minority instance until the desired level of equilibrium is reached between the classes. This adaptive and iterative nature of ADASYN allows it to tailor the synthetic sample generation to the specific characteristics of the dataset, resulting in improved model performance by mitigating the impact of class imbalance. In our study, we utilized ADASYN to address the imbalance in the training-set during a stratified 10-fold cross-validation process, with the aim of enhancing the model's performance.

3.5. Multilayer Perceptron

We propose the utilization of a Multilayer Perceptron (MLP) for predicting customer behavior patterns, comprising a single input layer, two or more hidden layers, and a sole output layer. The predictive model is trained through the backpropagation technique [41,42]. Each unit within a layer is intricately connected to all units in the preceding layer, as depicted in Figure 4. This interconnectedness is maintained throughout the model. In two consecutive tiers, every pair of units is associated with a weight. The subsequent step involves computing the net input by multiplying each input with its corresponding weight and aggregating the results. The activation function is subsequently applied to this net input in each hidden layer unit. To minimize the disparity between predicted and target values, the backpropagation approach is employed for weight adjustment, a process performed iteratively after each training cycle. By following this iterative process, the most optimal model for the test set is derived. A critical component in evaluating the efficacy of a prediction model is the employment of a loss function, especially in scenarios involving binary classification with two classes. Our research has a specific focus on this binary classification task. The cross-entropy loss function stands as a suitable choice for computation, serving as a measure to assess the accuracy of predicting the desired outcomes. The computation of the cross-entropy loss function is accomplished as

$$\text{Loss} = -\sum\nolimits_i \left(y'_i \log(y_i) + \left(1 - y'_i\right) log(1 - y_i)\right) \tag{3}$$

where y_i' is true probability and y_i is predicted probability value. This process was repeated iteratively to determine optimal weights, ultimately resulting in optimal predictions for the test dataset.

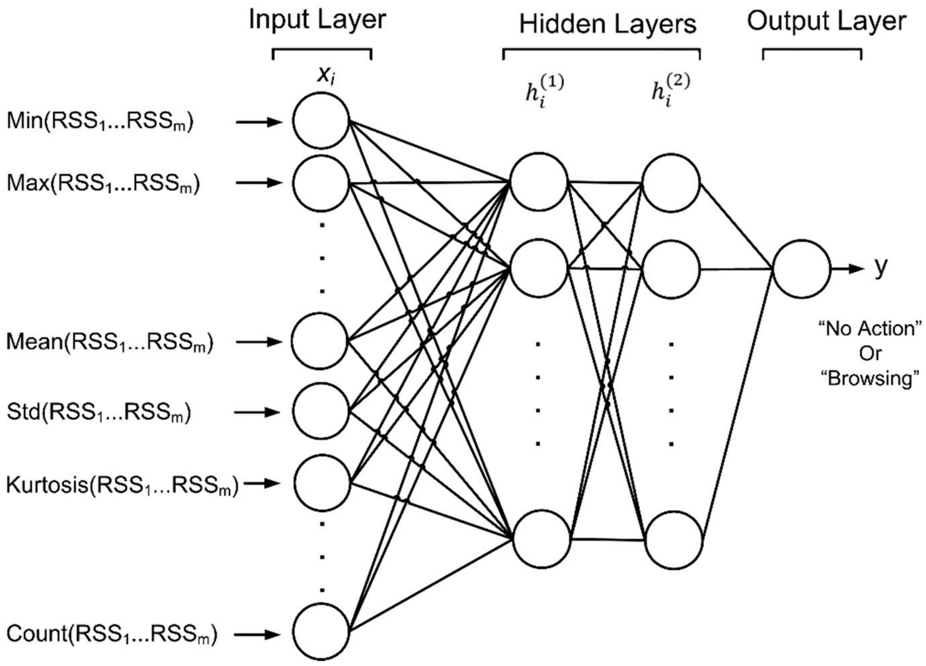


Figure 4. Proposed MLP model to predict customer behavior.

The time-domain characteristics extracted from the time-series dataset are transformed into a new structured dataset in tabular format. This tabular dataset is subsequently employed by a supervised machine learning model to acquire knowledge and make predictions for the testing dataset. MLPs can handle non-linear relationships and interactions among features, which are common in tabular data. Other machine learning models, including Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes (NB), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), and Random Forest (RF) were used to evaluate performance of customer shopping behaviors. These supervised machine learning models were found to be well-suited for extracting patterns from the tabular dataset, effectively discerning underlying relationships and trends within the structured data. Its capacity to identify and learn from these patterns allowed it to make accurate predictions and derive valuable insights when applied to the tabular dataset, as shown also in previous studies [15,17].

Machine learning algorithms were utilized for categorizing customer shopping behavior. These classification algorithms, ADASYN and iForest were executed in Python through the employment of Scikit-learn, Imbalanced-learn, and XGBoost libraries, using the default settings [43]. The assessment of these models was carried out using a stratified 10-fold cross-validation approach. Eventually, the effectiveness of the models was presented by examining metrics such as true positive (TP), true negative (TN), false positive (FP), and false negative (FN) rates [44]. Accurate classifications are represented by TP and TN, while FP and FN correspond to instances that were inaccurately categorized (see Table 1 for reference).

Table 1. Classifier model performance metrics.

Metric	Formula
Precision	$\frac{TP}{(TP+FP)}$
Recall/Sensitivity	$\frac{TP}{(TP+FN)}$
Specificity	$\frac{TN}{(TN+FP)}$
F1 score	$\frac{2\times(Precision\times Recall)}{(Precision+Recall)}$
Accuracy	$\frac{(TP+TN)}{(TP+TN+FP+FN)}$

4. Result and Discussion

This segment examines both the efficacy of the proposed model and the impact of outlier detection and data balancing methods on its performance. We present a comparison between our suggested approach and prior studies that employed RFID for detecting customer behavior. Ultimately, we also showcase a potential way to create a web-based customer behavior analysis by incorporating the suggested Multi-Layer Perceptron model.

4.1. Performance of Machine Learning Models

In this study, supervised machine learning methods were utilized to differentiate between various customer shopping behaviors. This differentiation was accomplished by using data from sensors gathered by an RFID device. Time domain features were extracted from the Received Signal Strength (RSS) of tagged products, and these were utilized as input attributes for classification models. Furthermore, the iForest Outlier Detection and ADSYN techniques were exclusively applied to the proposed Multilayer Perceptron (MLP) model.

Table 2 presents a comparison of different model performances, considering metrics such as accuracy, precision, specificity, recall, and f1 score. The machine learning models, including Logistic Regression (LR), K-Nearest Neighbor (KNN), Decision Tree (DT), Support Vector Machine (SVM), Naïve Bayes (NB), Adaptive Boosting (AdaBoost), eXtreme Gradient Boosting (XGBoost), and Random Forest (RF), were contrasted with the suggested MLP-based model for distinguishing customer shopping behaviors. The outcomes demonstrated that the proposed model surpassed the other models with improvements of up to 97.778% in accuracy, 98.008% in precision, 98.333% in specificity, 98.333% in recall, and 97.750% in the f1-score.

Table 2. Performance evaluation results.

Model	Accuracy	Precision	Specificity	Recall	F1 Score
LR	92.197 ± 8.389	93.472 ± 8.392	90.536 ± 9.968	90.536 ± 9.968	91.268 ± 9.610
KNN	95.606 ± 8.304	95.833 ± 8.690	94.786 ± 9.804	94.786 ± 9.804	95.079 ± 9.506
DT	89.470 ± 13.638	91.972 ± 10.861	88.929 ± 13.870	88.929 ± 13.870	88.290 ± 14.926
SVM	97.424 ± 3.939	97.750 ± 3.571	97.286 ± 4.211	97.286 ± 4.211	97.282 ± 4.154
NB	93.939 ± 8.906	94.806 ± 8.571	93.321 ± 9.335	93.321 ± 9.335	93.527 ± 9.413
AdaBoost	91.970 ± 13.035	92.817 ± 12.689	91.929 ± 13.233	91.929 ± 13.233	91.469 ± 13.824
XGBoost	91.970 ± 13.035	92.817 ± 12.689	91.929 ± 13.233	91.929 ± 13.233	91.469 ± 13.824
RF	92.879 ± 11.217	93.913 ± 10.074	93.179 ± 10.542	93.179 ± 10.542	92.577 ± 11.575
Proposed Model	97.778 ± 6.667	98.000 ± 6.000	98.333 ± 5.000	98.333 ± 5.000	97.750 ± 6.750

To assess the model's performance, we extended our evaluation to include the Receiver Operating Characteristic curve (ROC), a specialized metric for datasets with class imbalances [45]. The ROC curve helps to distinguish between false positives and false negatives, with an AUC (Area Under the Curve) value close to 1 indicating the best model performance [46]. Figure 5 showcases the ROC curve analysis for both the proposed and additional classification models we considered. Our results revealed that the suggested model achieved the highest AUC, reaching a value of 0.98.

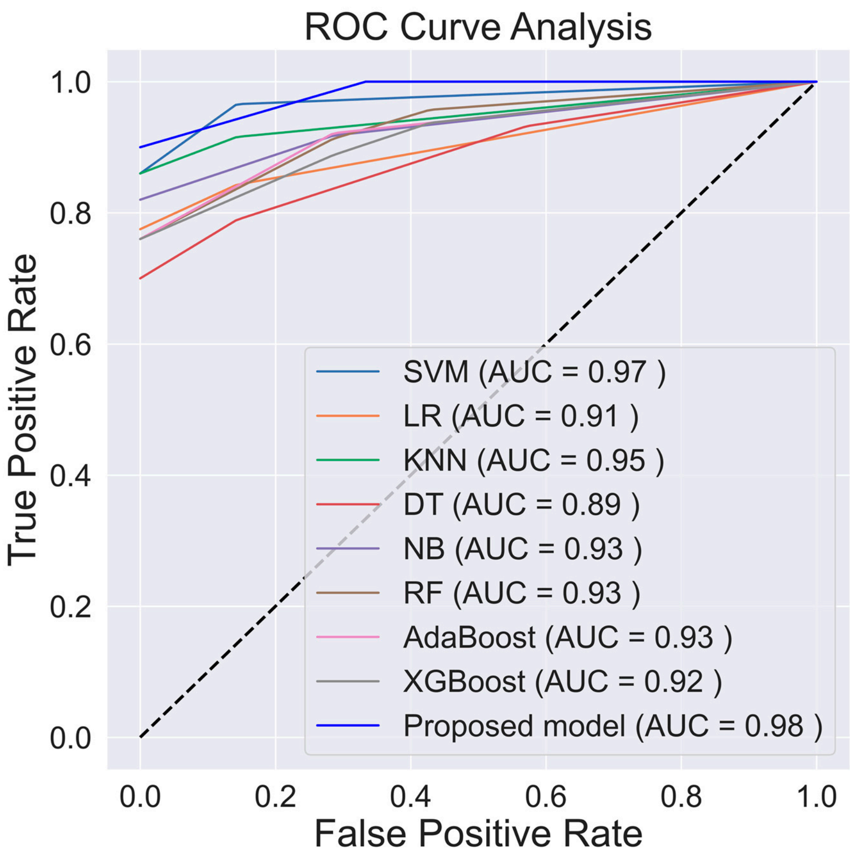


Figure 5. ROC analysis for the customer behavior prediction models.

The outcomes of the experiments demonstrate that the proposed MLP model can effectively identify customer browsing behavior on specific products with a notable level of accuracy. Nevertheless, it is crucial to grasp the interpretability of the predictive model and its real-world consequences. By employing the suggested model, managers gain an improved comprehension of customer browsing tendencies and their preferences for particular products. Retail managers can leverage these insights to enhance promotional strategies and offer pertinent product suggestions to customers. Moreover, managers have the opportunity to assess the arrangement of products in the retail store by eliminating less popular items from the layout and highlighting other products that have the potential to increase sales. Ultimately, as customer engagement and shopping quality rise, the bond between customers and the retailer strengthens, leading to a positive impact on the business's revenue.

4.2. Impact of Outlier Detection and Data Balancing Method on Model Performance

In this study, our primary objective was to explore the impact of incorporating iForest outlier detection and ADASYN data balancing on the accuracy of classification models. Our findings suggest that employing these techniques resulted in an improvement in model accuracy. Upon the implementation of iForest outlier detection, the classification models exhibited an average increase of 0.082% compared to machine learning models without iForest outlier detection. The integration of ADASYN with classification models led to an average enhancement of 0.808% compared to models without the ADASYN method. Notably, the simultaneous integration of both iForest and ADASYN into classification models resulted in an average increase of 2.14% compared to the original machine learning models. Further insights illustrating the effects of iForest and ADASYN on classification accuracy can be found in Figure 6. In summary, the incorporation of iForest and ADASYN into classification models holds the potential to significantly enhance overall model accuracy within our dataset.

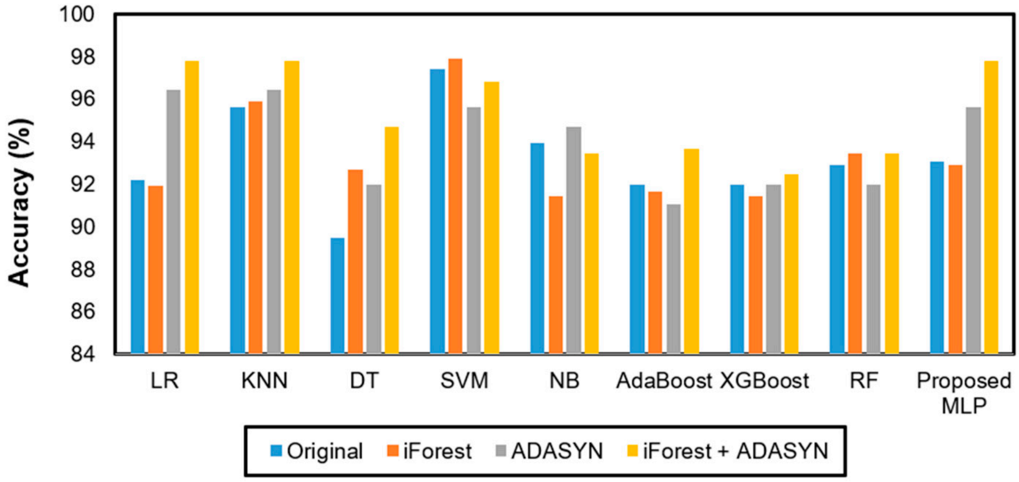


Figure 6. Impact of iForest and ADASYN on model prediction accuracy.

Nonetheless, employing synthetic data generation techniques carries the potential of generating data that are unrealistic or artificial, failing to accurately represent the inherent patterns in the actual data. Without careful design, these methods may also introduce bias, resulting in skewed outcomes and misleading model performance assessments. Furthermore, the process of generating synthetic data can be demanding in terms of computation and time, with the quality of the generated data greatly influenced by the selected generation method and parameters. To ensure that synthetic data enhance model training and generalization without unexpected complications, it is imperative to conduct analyses thorough validation and evaluation. Hence, in this study, we employed ADASYN to generate synthetic data that faithfully replicated the traits and distribution of the actual data, producing authentic artificial data that accurately represents the minority class. To validate the model, we also implemented stratified cross-validation, ensuring that the sample proportions for each class were maintained in each fold.

4.3. Comparison with Previous Studies

In this section, we conducted a comparative analysis between our study and prior research that focused on detecting customer activity through RFID technology. Table 3 outlines the distinctions between our findings and those of earlier investigations. This comparison considers a range of attributes, including input features, the machine learning algorithm employed, the achieved results, and the potential deployment of the model in real web or mobile environments.

In the context of detecting customer browsing behavior, Choi et al. [16] employed RFID tag occurrences and categorized them into distinct zones to comprehend customer browsing patterns. Zhou et al. [12] and Liu et al. [13], on the other hand, utilized phase readings as inputs to differentiate various types of customer behaviors, such as browsing and product rotation. Their prediction models exhibited a remarkable accuracy exceeding 93% [12] and an average precision of 89% [13]. Additionally, the identification of product movement types in retail stores relied on received signal strength (RSS) data and machine learning models. Hauser et al. [15] harnessed RSS from RFID tags and machine learning models to discern distinctive movement patterns inside the store. Meanwhile, Alfian et al. [17] employed RSS in conjunction with time-domain features to predict customer browsing activity within stores.

Table 3. Comparison of our study with previous work.

Authors	Input Feature	Model	Result	Model Deployment
[16]	Count (total RFID tag read occurrences)	K-Means clustering	The data from tag reads were grouped into multiple zones, with one of these zones representing customer browsing behavior.	Not reported
[12]	Phase shift	Gaussian model, ShopMiner	The model aimed to identify customer behaviors such as browsing and product interaction (turnaround), achieving an overall detection accuracy of more than 93%.	Not reported
[13]	RFID phase	Threshold based behavior recognition	The study attempted to identify customer behaviors, including browsing and product rotation, with an average precision of 89%.	Not reported
[15]	RSS with time-domain features	ANN	The model aimed to predict customer movement as they passed through the RFID gate, achieving an accuracy of 98.59%.	Not reported.
[17]	RSS with time-domain features	MLP + Extra Trees Feature Selection	The model's objective was to predict customer actions, achieving an accuracy of 97.00%, a precision of 96.67%, a recall of 97.50%, and an F1-score of 96.57%.	Not reported
Proposed Model	RSS with time-domain features	MLP + iForest outler detection + ADASYN data balancing	The model successfully predicted customer activity, achieving an accuracy of 97.778%, a precision of 98.008%, a specificity of 98.333%, a recall of 98.333%, and an F1-score of 97.750%.	Yes

It is worth noting that Table 3 does not serve as definitive proof of model performance, but rather offers a broad comparison that facilitates discussions about our proposed model and previous methodologies. Our study utilized a dataset initially provided by [17] and expanded it with an additional 8 RFID data readings. Furthermore, we incorporated iForest outlier detection and ADASYN to balance the training set, resulting in improved model performance as compared to previous study [17]. Finally, unlike prior studies, we deployed our trained model into web-based applications to empower managers with a comprehensive understanding of product popularity and customer interest.

4.4. Practical Applications

The primary objective of this study is to create a web-based system using machine learning. Previous studies have highlighted the system's usefulness in traceability system [47], inventory management [25], and disease prediction [48,49]. Furthermore, a prior investigation has demonstrated the integration of machine learning-based fall detection into an Internet of Things (IoT) device, exhibiting noteworthy performance results [50]. Our study aims to accurately predict customer shopping behavior, assisting in managerial decisions. By utilizing a machine learning model, it becomes feasible to distinguish between customer actions such as "browsing" and "no action". The web-based analysis of customer behavior was developed using PHP and a MySQL database on the server side. Python was used for the REST API and machine learning component. The predictive model was built using the Flask web framework and the Scikit-learn library on the server side. This model was employed to categorize types of customer activities. In Figure 7, the process is illustrated where sensor data (RSS) from RFID readers on store shelves are transmitted to the server side. The time-domain features were extracted from the RSS and the trained model was utilized to predict whether customers were "browsing" or taking "no action" with specific tagged products. The results were then presented to management through a web-based interface, depicted in Figure 8.

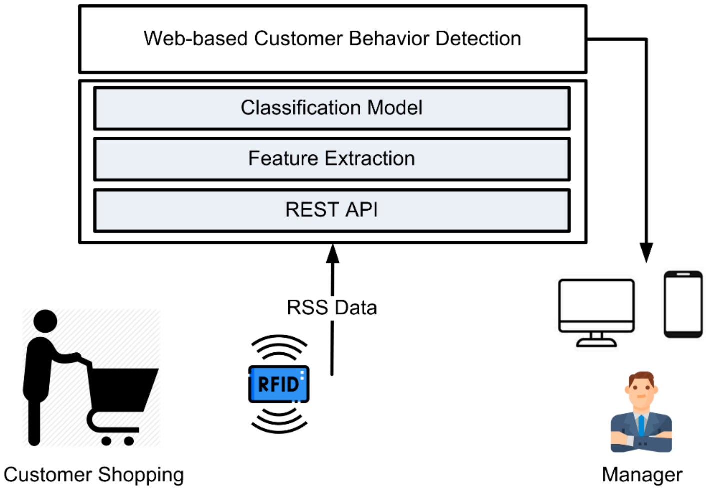
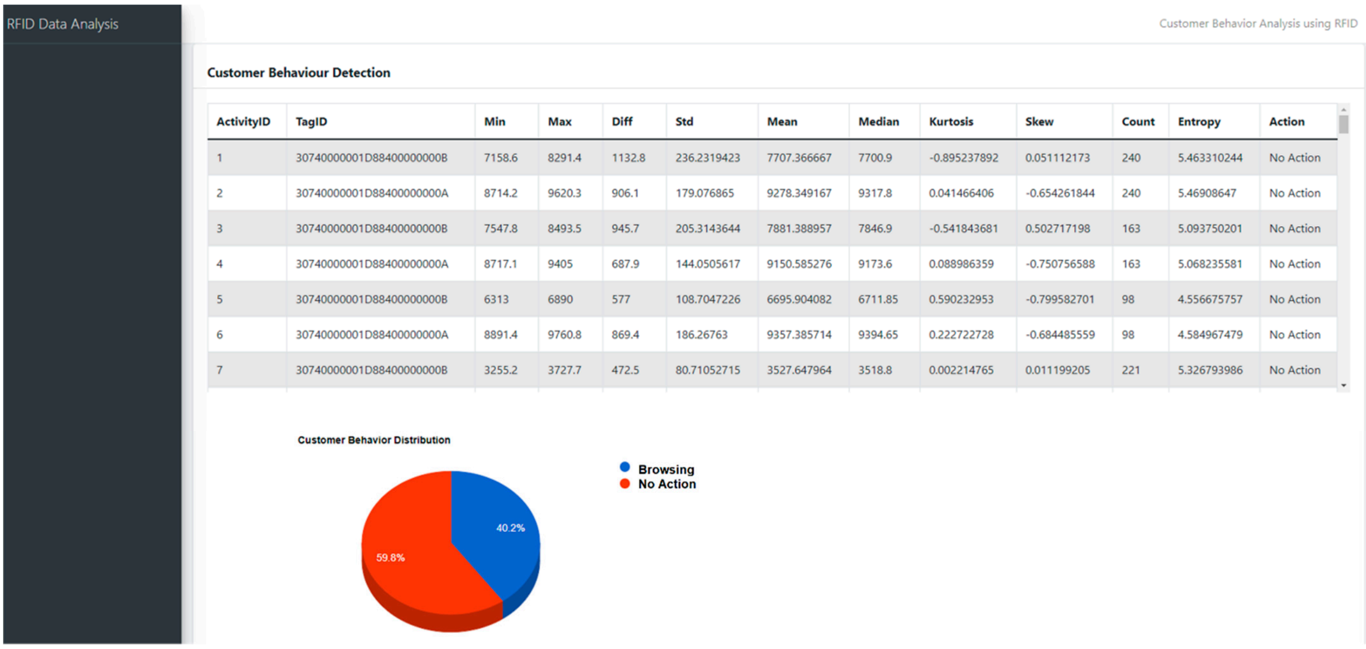


Figure 7. System Architecture of Web-based Customer Behavior Analysis.



ActivityID	TagID	Min	Max	Diff	Std	Mean	Median	Kurtosis	Skew	Count	Entropy	Action
1	30740000001D88400000000B	7158.6	8291.4	1132.8	236.2319423	7707.366667	7700.9	-0.895237892	0.051112173	240	5.463310244	No Action
2	30740000001D88400000000A	8714.2	9620.3	906.1	179.076865	9278.349167	9317.8	0.041466406	-0.654261844	240	5.46908647	No Action
3	30740000001D88400000000B	7547.8	8493.5	945.7	205.3143644	7881.388957	7846.9	-0.541843681	0.502717198	163	5.093750201	No Action
4	30740000001D88400000000A	8717.1	9405	687.9	144.0505617	9150.585276	9173.6	0.088986359	-0.750756588	163	5.068235581	No Action
5	30740000001D88400000000B	6313	6890	577	108.7047226	6695.904082	6711.85	0.590232953	-0.799582701	98	4.556675757	No Action
6	30740000001D88400000000A	8891.4	9760.8	869.4	186.26763	9357.385714	9394.65	0.222722728	-0.684485559	98	4.584967479	No Action
7	30740000001D88400000000B	3255.2	3727.7	472.5	80.71052715	3527.647964	3518.8	0.002214765	0.011199205	221	5.326793986	No Action

Figure 8. Dashboard of Web-based Customer Behavior Analysis.

Figure 7 illustrates the customers engagement with tagged products, indicating whether they are browsing or not interested. By categorizing and quantifying the frequency of customers engaging in browsing activities, managers can gain a comprehensive understanding of product popularity and customer interest. This data-driven approach allows managers to identify browsing patterns, thus can aid in optimizing the store layout and enhancing the overall shopping experience. With such insights, managers can tailor their strategies to effectively cater to customer preferences, ultimately leading to increased customer satisfaction and improved sales.

However, developing a web-based application comes with various potential challenges, including ensuring robust security, scalability to handle increased user loads, compatibility across diverse platforms, optimizing performance, creating a seamless user experience, complying with data privacy regulations, managing ongoing maintenance costs, and keeping the application up to date. In parallel, gathering user feedback is essential for improving the application. Users typically provide insights on usability, performance, feature requests, bug reports, security concerns, compatibility issues, content quality, and enhancement suggestions. This user feedback serves as a valuable source of information for enhancing the application's functionality, addressing issues, and aligning it with user needs and expectations.

5. Conclusions and Future Works

Retail store managers commonly participate in manual evaluations to comprehend customer behavior. However, this approach demands a substantial time investment. Conversely, the adoption of information technology streamlines managerial tasks, accelerating the decision-making process. Recent advancements in customer behavior analysis have arisen from the fusion of RFID technology and machine learning algorithms. In this study, the MLP model, in conjunction with iForest and ADASYN, effectively discerned customer activities using RFID sensors. Time-domain features were extracted from RFID tag RSS data, while machine learning models predicted customer engagement with specific tagged products, distinguishing between browsing and disinterest. Results highlighted the superior performance of the proposed model compared to alternatives such as LR, KNN, DT, SVM, NB, XGBoost, Adaboost, and RF, exhibiting enhancements of up to 97.778% in accuracy, 98.008% in precision, 98.333% in specificity, 98.333% in recall, and 97.750% in f1-score. This trained model integrated into a web-based system could receive RFID reader-derived RSS data to forecast customer behavior, furnishing valuable insights for managerial decision making, optimizing store layout, and enriching the overall shopping encounter.

In future research, it would be beneficial to delve into more complex real-world scenarios that involve a larger and more diverse dataset of customer shopping behaviors. Furthermore, the incorporation of varied time series feature-extraction methods, feature selection methods, and an expanded array of classification models could be considered in forthcoming studies.

Author Contributions: Conceptualization, G.A., N.L.F. and M.S.; Data curation, G.A., D.G.P.P., F.S. and M.S.; Formal analysis, D.G.P.P., N.L.F., F.T.D.A. and U.F.; Funding acquisition, M.S.; Investigation, F.M.H., R.A.N., Y.M.S., D.G.P.P., F.S., F.T.D.A. and U.F.; Methodology, G.A.; Project administration, M.Q.H.O., F.M.H., F.T.D.A. and D.T.N.; Resources, M.Q.H.O., F.M.H., R.A.N., Y.M.S., D.G.P.P. and F.S.; Software, G.A. and M.Q.H.O.; Supervision, Y.M.S., U.F. and D.T.N.; Validation, M.Q.H.O., F.M.H., R.A.N., Y.M.S., F.S., F.T.D.A. and D.T.N.; Visualization, R.A.N., N.L.F., U.F. and D.T.N.; Writing—original draft, G.A.; Writing—review and editing, N.L.F. and M.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by RTA Program Universitas Gadjah Mada with the Grant Number 5075/UN1.P.II/Dit-Lit/PT.01.01/2023.

Data Availability Statement: The data that support the findings of this study is publicly available at https://github.com/ganjar87/RFID_customer_behavior (accessed on 7 October 2023).

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Hawkins, D.I.; Mothersbaugh, D.L. *Consumer Behavior: Building Marketing Strategy*, 13th ed.; McGraw-Hill Education: New York, NY, USA, 2016; ISBN 978-1-259-23254-1.
2. Want, R. An Introduction to RFID Technology. *IEEE Pervasive Comput.* **2006**, *5*, 25–33. [CrossRef]
3. Ozguven, E.E.; Ozbay, K. An RFID-Based Inventory Management Framework for Emergency Relief Operations. *Transp. Res. Part C Emerg. Technol.* **2015**, *57*, 166–187. [CrossRef]
4. Lenko, F. Specifics of RFID Based Access Control Systems Used in Logistics Centers. *Transp. Res. Procedia* **2021**, *55*, 1613–1619. [CrossRef]
5. Oztekin, A.; Pajouh, F.M.; Delen, D.; Swim, L.K. An RFID Network Design Methodology for Asset Tracking in Healthcare. *Decis. Support Syst.* **2010**, *49*, 100–109. [CrossRef]
6. Unhelkar, B.; Joshi, S.; Sharma, M.; Prakash, S.; Mani, A.K.; Prasad, M. Enhancing Supply Chain Performance Using RFID Technology and Decision Support Systems in the Industry 4.0–A Systematic Literature Review. *Int. J. Inf. Manag. Data Insights* **2022**, *2*, 100084. [CrossRef]
7. Badi, R.; Momin, B. SISC: Sensor-based intelligent shopping cart. In Proceedings of the 2018 3rd International Conference for Convergence in Technology (I2CT), Pune, India, 6–8 April 2018; IEEE: New York, NY, USA, 2018; pp. 1–4.

8. Athauda, T.; Marin, J.C.L.; Lee, J.; Karmakar, N.C. Robust Low-Cost Passive UHF RFID Based Smart Shopping Trolley. *IEEE J. Radio Freq. Identif.* **2018**, *2*, 134–143. [CrossRef]
9. Pradhan, S.; Chai, E.; Sundaresan, K.; Rangarajan, S.; Qiu, L. Konark: A RFID based system for enhancing in-store shopping experience. In Proceedings of the 4th International on Workshop on Physical Analytics, Niagara Falls, NY, USA, 19 June 2017; ACM: New York, NY, USA, 2017; pp. 19–24.
10. Nakahara, T.; Yada, K. Analyzing Consumers' Shopping Behavior Using RFID Data and Pattern Mining. *Adv. Data Anal. Classif.* **2012**, *6*, 355–365. [CrossRef]
11. Shen, B.; Zheng, Q.; Li, X.; Xu, L. A Framework for Mining Actionable Navigation Patterns from In-Store RFID Datasets via Indoor Mapping. *Sensors* **2015**, *15*, 5344–5375. [CrossRef]
12. Zhou, Z.; Shangguan, L.; Zheng, X.; Yang, L.; Liu, Y. Design and Implementation of an RFID-Based Customer Shopping Behavior Mining System. *IEEE/ACM Trans. Netw.* **2017**, *25*, 2405–2418. [CrossRef]
13. Liu, W.; Wang, J.; Li, B.; Wang, L. Enabling Fine-Grained Shopping Behavior Information Acquisition with Dual RFID Tags. *IEEE Internet Things J.* **2020**, *7*, 7539–7549. [CrossRef]
14. Shangguan, L.; Zhou, Z.; Zheng, X.; Yang, L.; Liu, Y.; Han, J. ShopMiner: Mining customer shopping behavior in physical clothing stores with COTS RFID devices. In Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, Seoul, Republic of Korea, 1–4 November 2015; ACM: New York, NY, USA, 2015; pp. 113–125.
15. Hauser, M.; Günther, S.A.; Flath, C.M.; Thiesse, F. Towards Digital Transformation in Fashion Retailing: A Design-Oriented IS Research Study of Automated Checkout Systems. *Bus. Inf. Syst. Eng.* **2019**, *61*, 51–66. [CrossRef]
16. Choi, S.H.; Yang, Y.X.; Yang, B.; Cheung, H.H. Item-Level RFID for Enhancement of Customer Shopping Experience in Apparel Retail. *Comput. Ind.* **2015**, *71*, 10–23. [CrossRef]
17. Alfian, G.; Syafrudin, M.; Rhee, J.; Stasa, P.; Mulyanto, A.; Fatwanto, A. In-Store Customer Shopping Behavior Analysis by Utilizing RFID-Enabled Shelf and Multilayer Perceptron Model. *IOP Conf. Ser. Mater. Sci. Eng.* **2020**, *803*, 012022. [CrossRef]
18. Pasanisi, S.; Paiano, R. A Hybrid Information Mining Approach for Knowledge Discovery in Cardiovascular Disease (CVD). *Information* **2018**, *9*, 90. [CrossRef]
19. Javed, F.; Gilani, S.O.; Latif, S.; Waris, A.; Jamil, M.; Waqas, A. Predicting Risk of Antenatal Depression and Anxiety Using Multi-Layer Perceptrons and Support Vector Machines. *J. Pers. Med.* **2021**, *11*, 199. [CrossRef]
20. Ke, K.-C.; Huang, M.-S. Quality Prediction for Injection Molding by Using a Multilayer Perceptron Neural Network. *Polymers* **2020**, *12*, 1812. [CrossRef] [PubMed]
21. Ferreira, J.; Callou, G.; Josua, A.; Tutsch, D.; Maciel, P. An Artificial Neural Network Approach to Forecast the Environmental Impact of Data Centers. *Information* **2019**, *10*, 113. [CrossRef]
22. Ahmed, S. A Software Framework for Predicting the Maize Yield Using Modified Multi-Layer Perceptron. *Sustainability* **2023**, *15*, 3017. [CrossRef]
23. Ribeiro, D.; Matos, L.M.; Moreira, G.; Pilastri, A.; Cortez, P. Isolation Forests and Deep Autoencoders for Industrial Screw Tightening Anomaly Detection. *Computers* **2022**, *11*, 54. [CrossRef]
24. Chen, J.; Zhang, J.; Qian, R.; Yuan, J.; Ren, Y. An Anomaly Detection Method for Wireless Sensor Networks Based on the Improved Isolation Forest. *Appl. Sci.* **2023**, *13*, 702. [CrossRef]
25. Alfian, G.; Syafrudin, M.; Fitriyani, N.L.; Alam, S.; Pratomo, D.N.; Subekti, L.; Octava, M.Q.H.; Yulianingsih, N.D.; Atmaji, F.T.D.; Benes, F. Utilizing Random Forest with IForest-Based Outlier Detection and SMOTE to Detect Movement and Direction of RFID Tags. *Future Internet* **2023**, *15*, 103. [CrossRef]
26. Li, Z.; Wang, J.; Zhang, Z.; Jin, F.; Yang, J.; Sun, W.; Cao, Y. A Method Based on Improved IForest for Trunk Extraction and Denoising of Individual Street Trees. *Remote Sens.* **2022**, *15*, 115. [CrossRef]
27. Zeng, J.; Cao, B.; Tian, R. Quality Monitoring for Micro Resistance Spot Welding with Class-Imbalanced Data Based on Anomaly Detection. *Appl. Sci.* **2020**, *10*, 4204. [CrossRef]
28. Javeed, A.; Dallora, A.L.; Berglund, J.S.; Idrisoglu, A.; Ali, L.; Rauf, H.T.; Anderberg, P. Early Prediction of Dementia Using Feature Extraction Battery (FEB) and Optimized Support Vector Machine (SVM) for Classification. *Biomedicines* **2023**, *11*, 439. [CrossRef] [PubMed]
29. Askari, M.R.; Abdel-Latif, M.; Rashid, M.; Sevil, M.; Cinar, A. Detection and Classification of Unannounced Physical Activities and Acute Psychological Stress Events for Interventions in Diabetes Treatment. *Algorithms* **2022**, *15*, 352. [CrossRef]
30. Raihan, M.; Hassan, M.M.; Hasan, T.; Bulbul, A.A.-M.; Hasan, M.K.; Hossain, M.S.; Roy, D.S.; Awal, M.A. Development of a Smartphone-Based Expert System for COVID-19 Risk Prediction at Early Stage. *Bioengineering* **2022**, *9*, 281. [CrossRef]
31. Zakariah, M.; AlQahtani, S.A.; Al-Rakhami, M.S. Machine Learning-Based Adaptive Synthetic Sampling Technique for Intrusion Detection. *Appl. Sci.* **2023**, *13*, 6504. [CrossRef]
32. Cao, B.; Li, C.; Song, Y.; Qin, Y.; Chen, C. Network Intrusion Detection Model Based on CNN and GRU. *Appl. Sci.* **2022**, *12*, 4184. [CrossRef]
33. Fu, Y.; Du, Y.; Cao, Z.; Li, Q.; Xiang, W. A Deep Learning Model for Network Intrusion Detection with Imbalanced Data. *Electronics* **2022**, *11*, 898. [CrossRef]
34. Thaher, T.; Zaguia, A.; Al Azwari, S.; Mafarja, M.; Chantar, H.; Abuhamdah, A.; Turabieh, H.; Mirjalili, S.; Sheta, A. An Enhanced Evolutionary Student Performance Prediction Model Using Whale Optimization Algorithm Boosted with Sine-Cosine Mechanism. *Appl. Sci.* **2021**, *11*, 10237. [CrossRef]

35. Li, Q.; Song, Z. Ensemble-Learning-Based Prediction of Steel Bridge Deck Defect Condition. *Appl. Sci.* **2022**, *12*, 5442. [CrossRef]
36. Liu, F.T.; Ting, K.M.; Zhou, Z.-H. Isolation forest. In Proceedings of the 2008 Eighth IEEE International Conference on Data Mining, Pisa, Italy, 15–19 December 2008; IEEE: New York, NY, USA, 2008; pp. 413–422.
37. He, H.; Bai, Y.; Garcia, E.A.; Li, S. ADASYN: Adaptive synthetic sampling approach for imbalanced learning. In Proceedings of the 2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence), Hong Kong, 1–8 June 2008; IEEE: New York, NY, USA, 2008; pp. 1322–1328.
38. Alien Technology ALR-9900 Enterprise RFID Reader. Available online: http://www.alientechnology.com/wp-content/uploads/Alien-Technology-ALR-9900-Enterprise-RFID-Reader.pdf (accessed on 1 January 2023).
39. Ma, H.; Wang, Y.; Wang, K. Automatic Detection of False Positive RFID Readings Using Machine Learning Algorithms. *Expert Syst. Appl.* **2018**, *91*, 442–451. [CrossRef]
40. Alfian, G.; Syafrudin, M.; Yoon, B.; Rhee, J. False Positive RFID Detection Using Classification Models. *Appl. Sci.* **2019**, *9*, 1154. [CrossRef]
41. Rumelhart, D.E.; Hinton, G.E.; Williams, R.J. Learning Representations by Back-Propagating Errors. *Nature* **1986**, *323*, 533–536. [CrossRef]
42. Han, J.; Kamber, M.; Pei, J. *Data Mining*; Elsevier: Amsterdam, The Netherlands, 2012; ISBN 978-0-12-381479-1.
43. Pedregosa, F.; Varoquaux, G.; Gramfort, A.; Michel, V.; Thirion, B.; Grisel, O.; Blondel, M.; Müller, A.; Nothman, J.; Louppe, G.; et al. Scikit-Learn: Machine Learning in Python. *J. Mach. Learn. Res.* **2012**, *12*, 2825–2830. [CrossRef]
44. Sokolova, M.; Lapalme, G. A Systematic Analysis of Performance Measures for Classification Tasks. *Inf. Process. Manag.* **2009**, *45*, 427–437. [CrossRef]
45. Bradley, A.P. The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms. *Pattern Recognit.* **1997**, *30*, 1145–1159. [CrossRef]
46. Huang, J.; Ling, C.X. Using AUC and Accuracy in Evaluating Learning Algorithms. *IEEE Trans. Knowl. Data Eng.* **2005**, *17*, 299–310. [CrossRef]
47. Alfian, G.; Syafrudin, M.; Farooq, U.; Ma'arif, M.R.; Syaekhoni, M.A.; Fitriyani, N.L.; Lee, J.; Rhee, J. Improving Efficiency of RFID-Based Traceability System for Perishable Food by Utilizing IoT Sensors and Machine Learning Model. *Food Control* **2020**, *110*, 107016. [CrossRef]
48. Rau, H.-H.; Hsu, C.-Y.; Lin, Y.-A.; Atique, S.; Fuad, A.; Wei, L.-M.; Hsu, M.-H. Development of a Web-Based Liver Cancer Prediction Model for Type II Diabetes Patients by Using an Artificial Neural Network. *Comput. Methods Programs Biomed.* **2016**, *125*, 58–65. [CrossRef]
49. Ahmed, N.; Ahammed, R.; Islam, M.M.; Uddin, M.A.; Akhter, A.; Talukder, M.A.-A.; Paul, B.K. Machine Learning Based Diabetes Prediction and Development of Smart Web Application. *Int. J. Cogn. Comput. Eng.* **2021**, *2*, 229–241. [CrossRef]
50. Anceschi, E.; Bonifazi, G.; De Donato, M.C.; Corradini, E.; Ursino, D.; Virgili, L. SaveMeNow.AI: A machine learning based wearable device for fall detection in a workplace. In *Enabling AI Applications in Data Science*; Hassanien, A.-E., Taha, M.H.N., Khalifa, N.E.M., Eds.; Studies in Computational Intelligence; Springer International Publishing: Cham, Switzerland, 2021; Volume 911, pp. 493–514, ISBN 978-3-030-52066-3.