ZKG INTERNATIONAL ISSN: 2366-1313

Machine Learning Approaches for Predicting People's Abnormal Behaviors

[1]Shahawar Fatima,[2] Mir Ahmed Ali Khan Rizvi, [3]Abdul Muqtader, [4]Mohd Junaid Uddin

[1]Assistant professor, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

shahwar012@gmail.com

[2,3,4] BE Student, Dept of CSE-AI&ML, Lords Institute of Engineering and Technology, Hyd.

ahmedalikhanrizvi@gmail.com,abdulzain2002@gmail.com,mohdjunaiduddin58@gmail.com

Abstract: *Some errors in a few situations can also put humans at hazard, such as smoking in a gas station, which is why they ought to be checked. This article tries to discover an extremely good device mastering method to solve this prediction hassle. Data related to behavioral evaluation were accumulated, classes that protected smoking, telephone calls, and well known behavior. Experiments primarily based on many famous strategies had been completed, which includes help vector device (LSVM), kernel help vector device (KSVM), choice tree (DT) classifier, random forest classifier (RF), K-nearest friends (KNN). And K-Means clustering. In addition, the confusion matrix and the suggest square blunders (MSE) are used to determine the performance of each set of rules. Finally, principal element evaluation (PCA) confirmed the effects of the Nice set of rules. The outcomes display that the Random Forest Classifier (RF) achieves the overall overall performance and is capable of expect human conduct with 80- percent accuracy.*

Keywords—Machine Learning; Abnormal behaviorsprediction; Dimensionalreduction

I. INTRODUCTION

Today, human beings cope with their fitness, however there are many risky behaviors that may damage humans. They are very threatening in a few approaches. For example, speaking at the phone even as riding affects human's interest, which can cause vehicle accidents. Additionally, it is illegal to smoke in locations which include gasoline stations and stores due to the fact they are able to reason the range to explode even supposing it breaks. Avoiding sure bad behaviors may also last many people's lives. So the government has already handed many legal guidelines towards people's bad conduct and that they

ISSN: 2366-1313

need to be controlled in time. However, it isn't viable to verify such behavior in truth through private use. Fortunately, machine control and PC imaginative and prescient are a success and may be utilized by human beings. By reading the connection between information, computer systems can increase the potential to classify pix themselves. So if a few pictures of smoke and calls can be saved in laptop systems for studying purposes, they may be used to help influence terrible conduct.

Machine gaining knowledge of has evolved to a particular quantity in latest years [1-3]. In preceding research, some research have already tried to apply the gadget to advantage information in the region of computational questioning and prescience of human beings. By growing a constitutional neural community, the laptop can distinguish human conduct [4]. On a larger scale, Zhu et al. Also published a deep getting to know based set of rules to analyze student conduct throughout trying out [5]. Concerning smoking behavior, Zhang et al. Developed a device mastery manner inside the form of tree choice [6]. Their version achieves eighty four.11% accuracy with universal performance.

However, there's nonetheless a few research into predicting name behavior, specially using algorithms primarily based typically on machine studying strategies. For example, smoking, talking on mobile cellphone is difficult to locate in spite of the bare eye. The telephone may be too small and blocked by way of the human hand, causing greater stress. In [7], Zheng used device learning algorithms based on help vector gadget (SVM) further to constitutional neural community (CNN) to expect the behavior of human beings walking above and underneath, which changed into taken into consideration the maximum accurate at ninety three.5%. However, this article surely desires to assess the principle interest in gaining know-how approximately smoking detection algorithms and reporting behavior and dad and mom, which is the right answer.

The the rest of this record is divided into the following sections: Section 2 gives this statistics about garage and its troubleshooting techniques. Then the view of every magnificence and the effects of numerous machine studying algorithms can be entered into the element

We. Finally, popularity might be summarized in bankruptcy four.

II METHOD

A. Description and advance of documents

ISSN: 2366-1313

The materials used in this paper include 3 classes: smoke elegance, call for beauty, and normal class. For the best detection of smoking, this paper selects the data "Cigarette Smoker Detection" from Kaggle, which contains 805 images with different parameters [8]. For Elegance Hu, it has 1227 different images from TIANCHI DATA SET and 396 images from CSDN, its size is 3456 × 4608 [9, 10]. The normal class from Kaggle's "Person Face Data set", contains 10,000 images of 1024 × 1024 [11]. Examples are shown in Figures 1, 2 and 3.

Figure 1. Sample images in Smoking class

Figure 2. Images in Calling class from TIANCHI DATA SET

Figure 3. Sample images in Normal class

Processioning has 6 parts. First, d lib's get_frontal_face_detector feature is used to locate human faces in pics. So their behaviors consisting of speak me on the phone can be better detected. After that, the whole image is transformed to 64×sixty four. In the third step, the photograph is converted to gray by means of the cvt Color feature of cv2. In this sense, they've many similarities with machine gaining knowledge of. Then, with the intention to balance the statistics, around 700 photographs are first decided on for each elegance because the calling elegance only contains around seven hundred photos. Furthermore, this paper normalizes the data with the aid of dividing 255. Finally, the statistics is divided into train and take a look at, its education ratio is 0.Eight. Figures 5, 6 and seven display the complete records.

Figure 4. Prepossessed images in Smoking class

Figure 5. Prepossessed images in Calling class

B. Machine manage algorithms

This article used many famous mastering strategies, together with Support Vector Machine, Decision Tree, Random Forest, K-seed Community, K-Means. There are a

ISSN: 2366-1313

few commands about the algorithms, which may be found underneath.

A) Support Vector Machine (SVM): SVM is a supervised getting to know set of rules that may be used to clear up model or regression troubles. Its reason is to discover the aircraft within the N-dimensional area, capable of distribute critical statistics. To separate the two forms of facts from the commands, there are numerous unique forms of hyperplane to pick from. Therefore, SVM is designed to find the excellent hyperplane that has the maximum distance among registration factors to higher distribute the statistics factors.

SVM has several kernels to pick out from: "linear", "poly", "rbf", "sigmoid", "recomputed". C in SVM is a steady regular and Gamma is a key time period for "rbf", "poly", "sigmoid".

This paper chooses LSVM - "linear" kernel and KSVM - "rbf" kernel to take a look at the capacity of models in data type. In order to make no modifications feasible, those are the defaults, in which C is 1 and Gamma is "scale".

B) Decision tree: a decision tree is a non-parametric tracking algorithm that may be used to remedy issues of type or regression. It uses a version that looks like a tree that may show options and results. Each branch represents a very last examination and every page represents a collection of papers.

C) Random Forest: Random Forest is a set of policies that can be used to resolve issues of kind or regression by means of growing a couple of random bushes. For this type of paintings, the cease result of the random forest is the splendor that may be determined from the maximum tree. For the regression function, the final result is the prediction of each tree. The advantage of Random Forest is that it prevents the choice of bushes from over fitting to their colleges. The main parameter of Random Forest is "n_estimators", ie the wide variety of trees in the forest.

This article makes use of a version with n_estimators of 250 to remedy this hassle.

D) K-Neighbors Community (KNN): KNN is an unsupervised algorithm that acquires knowledge of rules that can be used to resolve type or regression problems. In facts category, the real content is assigned to the class that ranks highest among its closest pals. In the case of a regression, the end result is the average of the values of close strategic partners.

E) K-Means Clustering: K-Means is a grouping of policies so as to be used to get rid of clustering troubles. It calls a couple of groups to simply accept the exception and iterate to find the common of the poor

ISSN: 2366-1313

data points until it offers the first cost to categorize all the statistics gadgets. The important parameter of K-Means is "n_clusters", which is the wide variety of clusters into which the records ought to be divided. We could even must create an algorithm.

This article uses 3 categories and compares the outcomes with the unique traits of the information.

C. Need extra statistics

a) Principal Component Analysis (PCA): PCA is a dimensional reduction approach that can be utilized in studying research data and making model predictions. It seems that the primary product and its use need to alternate the statistics base, on occasion the usage of the most handy first essential additives and ignoring the rest.

B) Mean Square Error (MSE): MSE is often used as a loss characteristic to test the sum of squares of the error, that's the common difference between the visual variables and the proper cost. This paper uses MSE as a loss characteristic to calculate the accuracy of the modes.

C) Confusion Matrix: Confusion Matrix is a desk that indicates the general overall performance of the code. Each line represents the version in perfection and every line represents the version in predicted elegance. This article uses the confusion matrix functionality to evaluate in Python.

III RESULTANDDISCUSSION

Confusion matrix and results of the special algorithms (i.e. KSVM, LSVM, Decision Tree, Random forest, KNN and K- manner) are showed in Table I, Table II, Table III, Table IV, Table V and Table VI.

TABLE I. CONFUSION MATRIX OF KSVM

CONFUSION MATRIX OF KSVM			
Predict / Actual	Smoking	Calling	Normal
Smoking	87	40	5
Calling	66	87	4
Normal	6	3	128

TABLE II. CONFUSION MATRIX OF LSVM

CONFUSION MATRIX OF LSVM			
Predict / Actual	Smoking	Calling	Normal
Smoking	105	25	2
Calling	59	91	7
Normal	2	1	134

TABLE III. CONFUSION MATRIX OF DECISION TREE

CONFUSION MATRIX OF DECISION TREE			
Predict / Actual	Smoking	Calling	Normal
Smoking	94	31	7
Calling	48	92	17
Normal	15	6	116

TABLE IV. CONFUSION MATRIX OF RANDOM FOREST

CONFUSION MATRIX OF RANDOM FOREST			
Predict / Actual	Smoking	Calling	Normal
Smoking	97	33	2
Calling	32	122	3
Normal	5	2	130

TABLE V. CONFUSION MATRIX OF KNN

CONFUSION MATRIX OF KNN			
Predict / Actual	Smoking	Calling	Normal
Smoking	93	9	30
Calling	78	45	34
Normal	2	0	135

TABLE VI. CONFUSION MATRIX OF K-MEANS

CONFUSION MATRIX OF K-Means			
Predict / Actual	Smoking	Calling	Normal
Smoking	33	67	32
Calling	38	74	45
Normal	109	3	25

TABLE VII. RESULTS OF DIFFERENT ALGORITHMS

RESULTS OF DIFFERENT ALGORITHMS					
	Accuracy	MSE	Precision	Recall	F1-score
KSVM	0.71	0.369	0.72	0.71	0.71
LSVM	0.77	0.254	0.78	0.77	0.77
Decision Tree	0.71	0.446	0.71	0.71	0.71
Random Forest	0.82	0.230	0.82	0.82	0.82
KNN	0.64	0.585	0.69	0.64	0.60
K-Means	0.31	1.683	0.33	0.31	0.31

ISSN: 2366-1313

A. The reality compares

In terms of accuracy, the consequences in Table VII display that the column assist vector device, kernel help vector system, pruning tree and random woodland are correct. Among them, Random Forest could be very a success, with an accuracy of eighty two%. KNN is not true, whilst K-person is the worst, the first-rate with an accuracy of 31%.

B. Mean Square Error Comparison

In phrases of common errors, the result showed that linear support vector device, kernel support vector device, choice tree, random woodland, KNN carry out properly, in particular helps linear, simplest has an MSE of zero.254. In contrast, the K-Means group has an MSE of one.683, that is a very good end result.

C. Discussion

The effects in precision and mean rectangular errors are stable. Random Forest has a mainly top performance inside the undertaking that this text attempts to clarify, whilst K-Means is the worst.

The cause why K-like has a horrible result may also lie within the reality that it's far a non-supervised set of rules, ideal for fixing troubles in the school room where pics do no longer incorporate textual content written. However, renovation algorithms may be extra efficient because all pix are already well classified.

D. Analysis of key elements

Figure eight shows that the very last result of the Random Forest algorithm found within the framework of the principle of analysis, analyzed with separate evidence a part of the data set, which additionally exists into Figure nine. The parent shows the primary order of the Random Forest as soon as. The highest pix are categorized well.

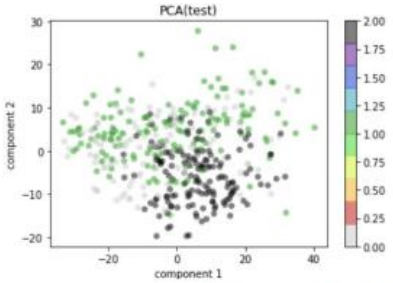

Figure 6. Visualization of the relationship between test_x and test_y.

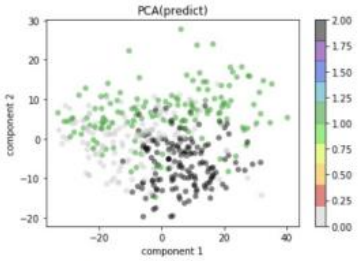

Figure 7. Visualization of the relationship between test_x and predicted_y by Random Forest.

IV CONCLUSION

The reason of looking is to denounce the terrible conduct of people who can placed the lives of others in danger by way of askingMachine studying algorithms. This article recognizes the overall performance of diverse system studying algorithms

ISSN: 2366-1313

along with Linear Support Vector Machine (LSVM), Kernel Support Vector Machine (KSVM), Decision Tree, Random Forest, K Community -seed (KNN) and K-Means clustering. The confusion matrix and suggest rectangular errors are achieved to assist select whether the model is accurate or now not. Additionally, the assessment supervisor discovers the result of the first rate set of rules. The analysis outcomes show that Random Forest is the first-rate layout for the trouble, even as K-Means clustering is the worst. In the future, programs that can be used to have a look at human behavior from a virtual digicam may be evolved and the Random Forest version may be similarly superior via the implementation of modifications.

REFERENCES

1. Q. Zhou, W. Lan, Y. Zhou and G. Mo, "Effectiveness Evaluation ofAnti-birdDevicesbasedonRandomForestAlgorithm," 20207thInternationalConferenceonInformation,Cybernetics,andComputationalSocialSystems(ICCSS),2020,pp.743-748.

2. Y. Guo, Y.Zhou, X. Hu andW. Cheng, "Researchon RecommendationofInsuranceProductsBasedonRandomForest," 2019InternationalConference on Machine Learning, Big Data and Business Intelligence(MLBDBI), 2019, pp. 308-311.

3. Y. Qiu, P. Chen, Z. Lin, Y. Yang, L. Zeng and Y.Fan, "ClusteringAnalysisforSilentTelecomCustomersBasedonK-means++," 2020IEEE4thInformationTechnology,Networking,ElectronicandAutomationControlConference(ITNEC),2020,pp.1023-1027.

4. S. Zheng, “Research on Algorithm of Pedestrain Attitude Estimation andRecognitionBasedonMachineLearning,”ShandongUniversity,2019,.DOI:10.27272/d.cnki.gshdu.2019.000463.

5. G.Zhu,X.Jiang,F.Xu,“Applicationofvideobehaviorandactionrecognitionbasedonmachinelearninginpaperlessassessment(inChinese),” Construction informational in China, 2019, vol. 10, pp. 56-57.

6. Y. Zhang, J. Liu, Z. Zhang and J. Huang. "Prediction of daily smokingbehavior based on decision tree machine learning algorithm." In 2019IEEE9thInternationalConferenceon ElectronicsInformationandEmergencyCommunication(ICEIEC),pp.330-333.IEEE,2019.

ISSN: 2366-1313

7.Y.Zheng,“ResearchonMachineLearning AlgorithmforHumanBehaviorRecognitio n,”WuhanUniversityofTechnology,2019. TheDOI: 10.27381/,dcnki.Gwlgu.2019.000606.“Ci garette r-smoker-detection”, kaggle,https://www.kaggle.com/dat asets/vitaminc/cigarette-smoker-detection,2019.

8. Prasadi Peddi and Dr. Akash Saxena (2015), "The Adoption of a Big Data and Extensive Multi-Labled Gradient Boosting System for Student Activity Analysis", International Journal of All Research Education and Scientific Methods (IJARESM), ISSN: 2455-6211, Volume 3, Issue 7, pp:68-73.

Prasadu Peddi (2016), Comparative study on cloud optimized resource and prediction using machine learning algorithm, ISSN: 2455-6300, volume 1, issue 3, pp: 88-94.