# Predicting age from hearing test results with machine learning

Eric Adams

Chapman University, CADS

CS 510 Reproducible Research

### Abstract

One of the important issues taken into consideration by most researchers is the proper Scrutiny of the ability to hear stimulus. In an attempt to help execute this experimentation, the research was undertaken to investigate, assess and compare models of distinct algorithmic classification methods in diagnosing the presence of hearing defects or not. In checking for the patterns of hearing defect occurrences over time and deciding on the best predictive models for the hearing defect analysis, data was collected from the Kaggle website as secondary source. A model developed for the forecasted number of reported defect cases was a compilation of several classification models stacked together with a learning base model of Quadratic Discriminant Analysis. The stacked classifier came up with an accuracy score of 93% better as compared to individual separate algorithms. Python software was used to analyze the data.

**Keywords:** Keyword; ML; Python, Age, Hearing Test, Physical Score

## INTODUCTION

Hearing aids still rely on oversimplified representations of auditory processing or hearing loss despite the complex and highly non-linear nature of sound processing in the human auditory system. Traditional means of identifying hearing defect technique has been ineffective in truly detecting and diagnosing complicated sensorineural impairments. Utilizing machine learning models can speed up the time-consuming process of developing and diagnosing the presence of hearing defects in humans. The benefits of applying standard algorithms help in identifying challenges individual faces in understanding based on age and physical scores on other factors. The dataset contains two features (age and physical score) and one categorical label (test result). It appears that class 1 indicates a good outcome (no hearing effect) and class 0 indicates a poor result (has a hearing defect) for the hearing test.

## METHODOLOGY

This chapter focuses mainly on the definition of machine learning, applications of machine learning, and its review for data analysis. Machine learning is the process of developing intelligent machines which can learn by itself. In, 1959 Arthur Samuel, a computer scientist at IBM and a pioneer in AI and computer gaming, coined the term machine learning. He defined it as a "field of study that gives computers the ability to learn without being explicitly programmed.

### Supervised Machine Learning

Machines are trained under the supervision of humans. These machines are trained using well "labeled data and the goal is to predict outcomes for the new variables. Datasets are referred to as labeled if the column and target variable are fully labeled. SML is very useful in risk assessment, image and object detection, fraud detection, etc. SML can be grouped into Regression and Classification. In the Regression algorithm, the output is expected to be continuous numeric data. Regressions algorithms include SLR, DTR, RFR, SVR, etc. In the Classification algorithm, the output is expected to be categorical data. This output can be grouped into Binary classification and Multi-Class Classification. Classification algorithms include KNN, DTC, Logistic Regressions, Naïve-Bayes, SVC, etc.

### Unsupervised Machine Learning

Machines are trained using unlabeled data without any supervision. This mechanism generally focuses on understanding relationships within datasets. Unlabeled data refers to a dataset with no column and target variables. Sample algorithms include but are not limited to Clustering Dimensionality Reductions such as PCA, K-means, DBSCAN, etc.

### Classification Algorithms

There are several classification algorithms that was considered in the project. Sample algorithms like Gradient Boost Classifier, logistic regression, SVC, Decision Tree Classifier, Random Forest Classifier, etc.

***Decision Trees*** A decision tree could be a flow-chart-like tree structure that uses a branching methodology to illustrate each outcome of a decision. every node inside the tree represents a check on a particular variable, and every branch is that the outcome of that test.

***Gradient Boost Classifier*** It is a tree-based supervised learning classification algorithm that uses an ensemble boosting technique for classification. It operates on combining weak learners to make strong learners. These weak learners are arranged sequentially to reduce the error (Boosting). It uses a Fully grown Decision Tree as a weak learner or base estimator with maximum depth from 8 to 32. Its base estimator is non-changeable. The learning happens by optimizing the loss functions (MSE, MAE, RMSE).

### Data Collection

The Hearing test data were gathered from online sources (Kaggle) with two feature variable (Age Physical Score) and a target variable (test result).

## DATA ANALYSIS

In this chapter, we shall look at the information that can be derived from the data collected and make inferences based on our outcome of results and then give conclusion and necessary recommendations.

### Data Representation

Online sources were used to gather the data for the hearing test study, which determines if a patient has hearing defect or not. NumPy, Pandas, Seaborn, Matplotlib, and other Python libraries were imported to clean, analyze, preprocess, and assess the data. The dataset was loaded into Jupyter Notebook, and information about its characteristics (such as its shape and description) was retrieved. The dataset was also examined for balance and missing values.

| Features | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| age | 5000.0 | 51.60900 | 11.28700sni1 | 18.0 | 143.0 | 51.0 | 60.0 | 90.0 |
| Physical score | 5000.0 | 32.76026 | 8.169802 | 0.0 | 26.7 | 35.3 | 38.9 | 50.0 |
| Test result | 5000.0 | 0.60000 | 0.489947 | 0.0 | 0.0 | 1.0 | 1.0 | 1.0 |

**Figure 1** Data Summary

From Figure 1, the summary statistics of 5000 observations were conducted for the experiment. The standard deviation of 11.287001 is an indication of the large dispersion of hearing detections in Age about the mean. Age has the highest mean value of 51.0 indicating a somewhat strong presence and effect of hearing existence. The maximum and minimum values depict the highest and lowest swings of recorded hearing defect cases respectively. There are no missing values present in the dataset. The records initially indicated an imbalance structure of "hearing defect": 3000 and "no hearing defect": 2000, however; using Fix Imbalance Methods like SMOTE to rebuild the "Sudo" dataset to balance the dataset. Hence, the dataset shows balanced hearing counts of 3000 test results of the hearing diagnosis.

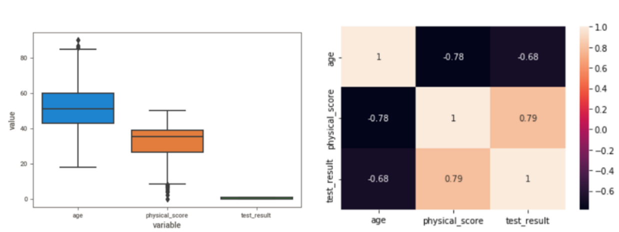


**Figure 2** Data Summary

***Preliminary Analysis II - Correlation Outlier Detection*** According to Figure 2, the correlation between Age with the test result is 68% (moderately positive) whiles the correlation between Physical Score and with test result is 79% (moderately Negative). Based on Figure 2, a boxplot for graphically demonstrating the locality and outlier detection indicates that there is an outlier. However, the outlier situation was fixed before the modeling development. The distribution of the dataset shows the normal distribution.

### Independent Variable Interactions

From Figure 5, the functionality package of "pair plot" in seaborn (visualization tool) enables the combinatorial interaction between two independent variables. The observed figure shows mild overlapping in its distribution, and this indicates the possibility of using tree base algorithms for its analysis due to its non-linearity behavior. However, tree base algorithm renders computational cost. If the overlapping intensity is wild, a KNN algorithm would have been the best option since it uses the concept of Euclidean/Manhattan distances to find the similarity of data points for its classification.

### Model Building

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC | TT (Sec) |
|---|---|---|---|---|---|---|---|---|---|
| qda | Quadratic Discriminant Analysis | 0.9325 | 0.9763 | 0.9627 | 0.9270 | 0.9443 | 0.8586 | 0.8602 | 0.037 |
| gbc | Gradient Boosting Classifier | 0.9257 | 0.9720 | 0.9621 | 0.9175 | 0.9391 | 0.8440 | 0.8462 | 0.344 |
| ada | Ada Boost Classifier | 0.9221 | 0.9628 | 0.9633 | 0.9111 | 0.9363 | 0.8362 | 0.8388 | 0.335 |
| lda | Linear Discriminant Analysis | 0.9221 | 0.9670 | 0.9783 | 0.8997 | 0.9373 | 0.8350 | 0.8405 | 0.037 |
| knn | K Neighbors Classifier | 0.9218 | 0.9541 | 0.9561 | 0.9164 | 0.9356 | 0.8360 | 0.8377 | 0.061 |
| ridge | Ridge Classifier | 0.9218 | 0.0000 | 0.9783 | 0.8992 | 0.9370 | 0.8342 | 0.8398 | 0.029 |
| lightgbm | Light Gradient Boosting Machine | 0.9218 | 0.9676 | 0.9561 | 0.9163 | 0.9356 | 0.8360 | 0.8377 | 0.233 |
| lr | Logistic Regression | 0.9214 | 0.9669 | 0.9495 | 0.9212 | 0.9349 | 0.8357 | 0.8369 | 0.636 |
| nb | Naive Bayes | 0.9193 | 0.9703 | 0.9363 | 0.9288 | 0.9323 | 0.8322 | 0.8328 | 0.022 |
| rf | Random Forest Classifier | 0.9071 | 0.9580 | 0.9314 | 0.9143 | 0.9226 | 0.8065 | 0.8074 | 0.671 |
| et | Extra Trees Classifier | 0.9021 | 0.9335 | 0.9248 | 0.9122 | 0.9182 | 0.7963 | 0.7972 | 0.467 |
| dt | Decision Tree Classifier | 0.8907 | 0.8949 | 0.9020 | 0.9135 | 0.9074 | 0.7740 | 0.7750 | 0.033 |
| svm | SVM - Linear Kernel | 0.8856 | 0.0000 | 0.9038 | 0.9106 | 0.9018 | 0.7640 | 0.7760 | 0.047 |

**Figure 3** Model Comparisons Without Hyperparameter Tuning

| Fold | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC |
|---|---|---|---|---|---|---|---|
| 0 | 0.9393 | 0.9686 | 0.9581 | 0.9412 | 0.9496 | 0.8733 | 0.8736 |
| 1 | 0.9393 | 0.9790 | 0.9581 | 0.9412 | 0.9496 | 0.8733 | 0.8736 |
| 2 | 0.9286 | 0.9833 | 0.9581 | 0.9249 | 0.9412 | 0.8503 | 0.8512 |
| 3 | 0.9607 | 0.9897 | 0.9759 | 0.9586 | 0.9672 | 0.9183 | 0.9185 |
| 4 | 0.9179 | 0.9776 | 0.9337 | 0.9281 | 0.9309 | 0.8296 | 0.8296 |
| 5 | 0.9143 | 0.9563 | 0.9759 | 0.8901 | 0.9310 | 0.8185 | 0.8245 |
| 6 | 0.9000 | 0.9623 | 0.9458 | 0.8920 | 0.9181 | 0.7900 | 0.7922 |
| 7 | 0.9357 | 0.9775 | 0.9518 | 0.9405 | 0.9461 | 0.8665 | 0.8666 |
| 8 | 0.9500 | 0.9839 | 0.9578 | 0.9578 | 0.9578 | 0.8964 | 0.8964 |
| 9 | 0.9140 | 0.9843 | 0.9217 | 0.9329 | 0.9273 | 0.8220 | 0.8221 |
| Mean | 0.9300 | 0.9763 | 0.9537 | 0.9307 | 0.9419 | 0.8538 | 0.8548 |
| Std | 0.0176 | 0.0101 | 0.0160 | 0.0224 | 0.0143 | 0.0372 | 0.0363 |

**Figure 4** Model Comparisons With Hyperparameter Tuning

Based on Figure 3 Figure 4, the output result of various algorithms is displayed. It can be identified that the quadratic discriminant analysis demonstrated outstanding metric performance with a high accuracy score value of 93.25% at 0.026 secs. This algorithmic metric performance without hyperparameter tuning was then followed by GBC, ADA, LDA, KNN, and Ridge respectively, etc. Nevertheless, from table 3.3.1(a), since the top five black box algorithms had similarly great metric performance with less time of iterations, a stacking classifier with hyperparameter tuning was considered (table 3.3.1(b)). Linear Discriminant Analysis has considered the main estimator while the remaining four algorithms (gbc, ada, lda, and knn) were considered base learners.

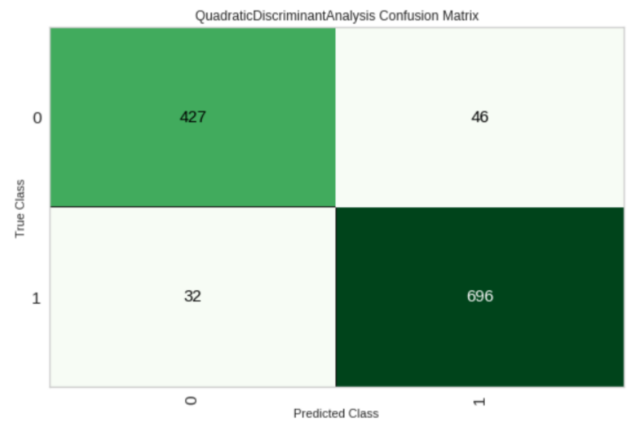


**Figure 5** Confusion Matrix


**Figure 6** Class Boundary

***Confusion Matrix Decision Boundary*** Figure 5, illustrates a special kind of contingency table with two dimensions ("actual" and "predicted"). The resulting analysis shows 78 misclassifications arising from both type I and type II errors. A totality value of 427 696 was correctly classified representing TP and FN respectively. Based on Figure 3.3.2(b), shows the occurrence of misclassification by assigning individual from " '0' : no hearing defect group" to a difference category of " '1' : individual with hearing defect" and vice versa.

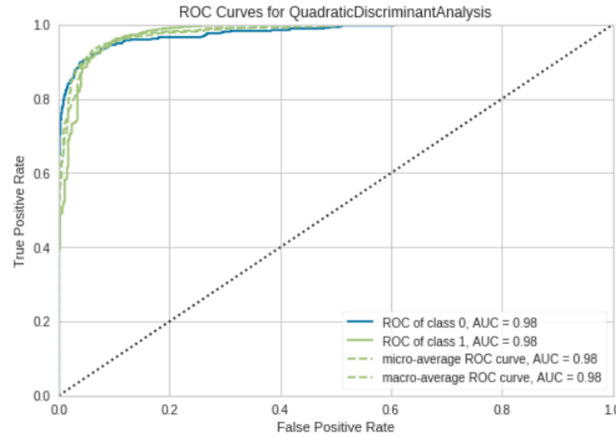


**Figure 7** ROAUC Curve

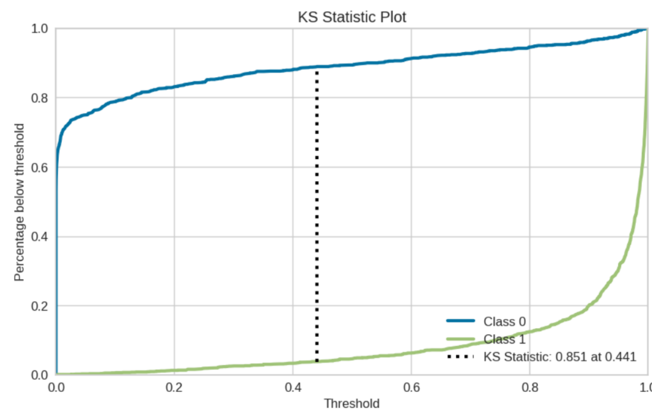


**Figure 8** Statistical Hypothesis

## Statistical Report AND ROAUC

Figure 7 Figure 8 represents the receiving operating characteristics curve performance measurement at various threshold settings. Upon using the stack classifier algorithm, the area under the curve shows a metric performance score of 98% which is greater than the threshold set value (50%). Hence, this indicates a great sign of a good model. The statistical plot shows a strong test statistic, and this indicates a statistical significance in our variable.

### Test Data Prediction

Below is the illustration of Prediction in Figure 9

## RESULT, CONCLUSION RECOMMENDATION

Based on the results obtained from this research work, the presence of varied age and physical score measurements can be used in predicting or determining the presence of a hearing defect in a patient. From the results displayed above, it can be concluded that Stacking Classifier with Quadratic Discriminant Analysis as the main estimator coupled with the other four (4) base learners came up with a great accuracy score for this project work. A feature importance extraction is necessary to be investigated on the QDA for further statistical analysis in determining the relative recurrence of hearing loss spreads on a particular feature based on the feature important extraction of the model. The advanced

| | Model | Accuracy | AUC | Recall | Prec. | F1 | Kappa | MCC |
|---|---|---|---|---|---|---|---|---|
| 0 | Quadratic Discriminant Analysis | 0.917 | 0.9709 | 0.9425 | 0.9228 | 0.9326 | 0.8247 | 0.825 |

| | age | physical_score | test_result | Label | Score |
|---|---|---|---|---|---|
| 0 | 39 | 42.1 | 1 | 1 | 0.9977 |
| 1 | 45 | 39.8 | 1 | 1 | 0.9912 |
| 2 | 65 | 20.8 | 0 | 0 | 1.0000 |
| 3 | 43 | 39.1 | 1 | 1 | 0.9939 |
| 4 | 68 | 23.3 | 0 | 0 | 1.0000 |
| 5 | 24 | 46.9 | 1 | 1 | 0.9999 |
| 6 | 42 | 39.3 | 1 | 1 | 0.9951 |
| 7 | 66 | 20.7 | 0 | 0 | 1.0000 |
| 8 | 52 | 35.4 | 0 | 1 | 0.9287 |
| 9 | 69 | 17.6 | 0 | 0 | 1.0000 |

**Figure 9** Test Data Prediction

study of black box algorithms about explainability and interpretability could help reduce the effect of computational cost. It should be noted that practically all the research discussed here ran validation tests to gauge how well their learning algorithms performed. Intuitively, their internal mechanism divides the initial datasets into subsets using well-known evaluation procedures. As previously indicated, significant and independent features that could lead to greater validation are necessary to achieve correct findings for their prediction models. These investigations included internal and external validation to enable the extraction of more precise and trustworthy predictions while minimizing any bias. The small number of data samples is one of the most typical constraints identified in the studies analyzed in this study. The size of the training datasets must be sufficient, which is a fundamental criterion for employing classification systems to simulate dysfunctional hearing behavior. An adequate division into training and testing sets is to be made possible by a relatively big dataset, which results in good estimator validation. A limited training sample compared to the dimensionality of the data might cause misclassifications, and the estimators can create models that are unstable and biased. It goes without saying that a larger patient population used to forecast patients' survival can improve the predictive model's generalizability. Even though correlation as a base tool for determining the relationship of variable interactions was used, however, an advanced study on "causal inference" could vividly aid in identifying the true parametric variable that could cause a change in the ecosystem of the dysfunctionality of hearing existence.

## REFERENCE

Le Goallec, Alan, et al. "Predicting age from hearing test results with machine learning reveals the genetic and environmental factors underlying accelerated auditory aging." medRxiv (2021).

Bing, D., et al. "Predicting the hearing outcome in sudden sensorineural hearing loss via machine learning models." Clinical Otolaryngology 43.3 (2018): 868-874.