

AIS-301

**Speech therapy classifier**

By: Nadeen Farid, Nadine Walid, Nouran Hady, Sama Okasha.

Submitted to: PROF.Ghada

Date: 31/12/2022

TABLE OF CONTENTS

Problem Statement………………………………………………………………………………...2

Related work…………………………………………………………………………………...….3

Model Architecture……………………………………………………………………………….4

Evaluation results...……………………………………………………………………………….8

Evaluation strategies outputs…………………………………………………………………….10

**Problem statement:**

We had a real dataset about 500 cases. The dataset is about patients having problems in articulation, mentally, and in concentration and hyperactivity. The original dataset was having over 14 class prediction ('Problem Behavior', 'ASD', 'Stuttering', 'DLD', 'PROBLEM IN ARTICULATION', 'H.H', 'Stuttering Problem Behavior','MR', 'ADHD', 'Hyperactive', 'Autizm', 'Echollia', 'DLD-MR'). But we managed to decrease the number of predicted classes to 8 classes by putting the more specified diagnosis under the umbrella of the more general diagnosis. The predicted classes are 8 ('Problem Behavior', 'Autizm', 'Stuttering', 'DLD', 'PROBLEM IN ARTICULATION', 'MR', 'ADHD', 'Hyperactive'). We had 26 features; the features are the features that we can measure and diagnose from them the problems we are trying to classify. The features are mainly classified to (evolution history of the case, evolution in language, behavior features). Our problem is supervised classification problem.

**Related work:**

We made a data analysis work on our dataset. We extracted the features from 34 features to 25 features. We also made exploratory data analysis and graphs to understand the distribution of the dataset. We filled the null data by 0 as the features are all discrete either 0 or 1. And the age column we filled it by median. And for the null values in the diagnosis column we removed the null values.

A picture containing graphical user interface

Description automatically generated

Chart, box and whisker chart

Description automatically generated

**Model Architecture:**

SVM model:

Support vector machine is a model created to Supervised Learning algorithms, which is used for Classification as well as Regression problems. The SVM intuition to decrease the error and increase the margin line.



• Max-Margin Classifier: Formalize the notion of the best linear separator

• Lagrangian Multipliers: Way to convert a constrained optimization problem to one that is

easier to solve

• Kernels: Projecting data into higher-dimensional space makes it linearly

separable

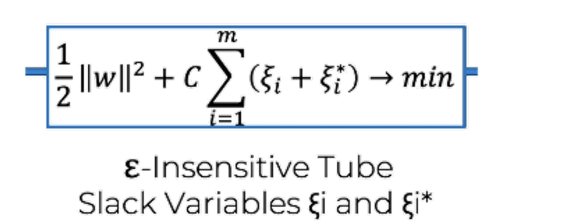
• Complexity: Depends only on the number of training examples, not on

dimensionality of the kernel space!

• Support Vector Machine (SVM) finds an optimal\* solution. Maximizes the distance between the hyperplane and the “difficult points” close to decision boundary

One intuition: if there are no points near the decision surface, then there are no very uncertain classification decisions

Least square sum error equation:



**Naive bayes Model:**

It is a classification technique based on Bayes theorem with an assumption of independence among features.

As in our project Speech therapy classifier, we try to find out the prediction whether the patient has ADHD, Scattering, and a lot of different diagnosis. Thus to find that it will be helpful and good choice to use bayes theorem to predict it. Thus after assuming that all your features are independent we use Naïve bayes theorem to predict. For Example:

We want to detect the diagnose. Thus, we put it

Thus, If we have all the numbers given we can easily predict the diagnose by placing the word diagnose with the specified one we want to predict.

**Note: Diagnose in the example should be replaced with the specified diagnose you want to predict.**

**Logistic regression:**

It is the best to describe our model since it is a classification model. The right one it accepts it with one and the wrong it counts it as zero. So, for example we want to predict some diagnosis and some syndrome we are going to predict it as the false counts as zero and the true counts as one. Since we have many diagnoses that we want, they should be replaced one after another.

**Decision Tree Classification Algorithm:**

Decision Tree Classification Algorithm

What is Decision Tree?

Decision Tree is a **Supervised learning technique** used for solving Classification problems.

Decision Tree Terminologies

* Root node / Leaf node
* Splitting
* Sub tree
* Pruning
* Parent/child node

**How does the Decision Tree algorithm Work?**

* **Step-1:** Begin the tree with the root node, which contains the complete dataset.
* **Step-2:** Find the best attribute in the dataset using Attribute Selection Measure**.**
* **Step-3:** Divide into subsets that contain possible values for the best attributes.
* **Step-4:** Generate the decision tree node, which contains the best attribute.
* **Step-5:** Recursively make new decision trees



**Information Gain**

Information gain is the measurement of changes in entropy after the segmentation of a dataset based on an attribute. And It calculates how much information a feature provides us about a class.

**Information Gain= Entropy(S)- [(Weighted Avg) \*Entropy (each feature)**

**Entropy:** Entropy is a metric to measure the impurity in each given attribute. It specifies randomness in data. Entropy can be calculated as:

Entropy(s)= -P(yes)log2 P(yes)- P(no) log2 P(no)

**S= Total number of samples** **P(yes)= probability of yes** **P(no)= probability of no**

**Evaluation results:**

**Svm confusion matrix**

**A screenshot of a computer

Description automatically generated with medium confidence**

**Naïve Bayes:**

**A screenshot of a computer

Description automatically generated with low confidence**

**Logistic regression:**

**A screenshot of a computer

Description automatically generated with low confidence**

**DECSION TREE:**

**Graphical user interface, text, application

Description automatically generated**

**Knn:**

**A picture containing text, black, electronics, keyboard

Description automatically generated**

**Evaluation strategies outputs:**

**Svm**

**Table

Description automatically generated**

**Naïve Bayes**

**Table

Description automatically generated**

**Logistic regression**

**Table

Description automatically generated**

**Decision tree**

**Table

Description automatically generated**

**Knn**

**Table

Description automatically generated**