

# Uso de algoritmos de aprendizado de máquina para predição de resultado do exame de cardiotocografia.

1<sup>st</sup> Anne Caroline Silva *IFES*

Serra, Brasil

carolinesilva4@hotmail.com

2<sup>nd</sup> Marcelo Passamai Mendes *IFES*

Serra, Brasil

marceloestudoifes@gmail.com

**Abstract**—A cardiotocografia é um exame realizado antes ou durante o trabalho de parto que monitora os dados de frequência cardíaca e contrações uterinas a fim de prevenir e diagnosticar o bem-estar fetal. Neste estudo, são utilizados três algoritmos de aprendizado de máquina - KNN, Árvore de Decisão e SVM - para classificação dos resultados que podem ser: normal, atípico/indeterminado e anormal. Para verificar a eficiência de cada algoritmo foram calculadas as seguintes métricas: acurácia.

**Index Terms**—Aprendizado de máquina, Cardiotocografia, Patologia

## I. INTRODUÇÃO

Um relatório da Organização Mundial de Saúde [?] revelou que cerca de 30 milhões de recém nascidos estão sob risco de vida, seja de morte, doenças ou incapacidade e que com cuidados básicos, de qualidade, cerca de 1.7 milhões de mortes poderiam ser evitadas.

Uma forma de avaliar e monitorar a saúde do feto é através da realização de um exame simples chamado cardiotocografia (CTG). Esse exame é indicado nas últimas semanas de gravidez, durante o parto ou quando a mulher apresenta alguma condição de saúde que eleve o risco da gravidez, tais como: diabetes gestacional, hipertensão arterial descontrolada, pré-eclâmpsia, anemia grave, alterações cardíacas, renais ou pulmonares, alterações no processo de coagulação sanguínea, infecções, idade avançada da mulher ou pouca idade.

Segundo Mello, 2019 [1], o CTG é feito através de cintos com sensores que são colocados sobre a barriga da gestante e em direção ao útero que captam os batimentos cardíacos do feto e a frequência e intensidade das contrações. Esses sensores emitem um sinal sonoro com frequência entre 500 e 1000 Hz e esperam uma reação do feto. O resultado do exame pode indicar se o feto está em sofrimento (quando há insuficiência na oxigenação cerebral). A partir desse resultado, o obstetra pode avaliar o melhor momento e forma de realizar o parto.

A baixa oxigenação pode ser ocasionada por: um resultado de posição do feto, problemas na placenta ou cordão umbilical enrolado no pescoço do bebê. Os sinais que indicam sofrimento fetal são:

- Taquicardia fetal (frequência cardíaca acima de 160 batimentos por minuto);

- Bradicardia fetal (frequência cardíaca abaixo de 110 batimentos por minuto);
- Quedas significativas na frequência cardíaca;
- Desacelerações na frequência cardíaca após as contrações.

O Objeto de estudo foi utilizado um *dataset* público, com 2126 registros com 23 atributos com resultados multicategóricos: 1 normal, 2 suspeito de patologia e 3 patológico.

## II. METODOLOGIA

Foi proposto o uso de 3 algoritmos clássicos: *knn*, *Support Vector Machine* e árvore de decisão.

### A. K-Nearest Neighbors

Segundo Vaz[2], o algoritmo *K-Nearest Neighbors*(KNN) dado um novo ponto, ou seja uma predição, o algoritmo calcula a distância desse ponto com todos os pontos conhecidos, dado um K que representa a quantidade de pontos próximos, em outras palavras *Nearest Neighbors*, o novo ponto é classificado como a classe mais frequente entre os *K-Nearest Neighbors*.

Uma observação importante, pelo motivo de boas práticas, esse número de K não deve ser par, para evitar empate de classificação.

Os parâmetros passados a função são a quantidade de vizinhos(K) e algoritmo de cálculo de distância que por padrão é a distância Euclidiana dada por:

$$\sqrt{((x_1 - x_2)^2 + (y_1 - y_2)^2)}$$

### B. Decision Tree

Segundo Junior[3], as árvores de decisão são uma estrutura baseada em grafos com a particularidade de serem: acíclico e conexo. As árvores possuem as seguintes características: um nó chamado de raiz onde é o início da árvore, nós, vértices e nós folha(nós sem adjacentes). Porém as árvores de decisão tem um detalhe a parte que cada nó que não é um nó folha pode ser representado por uma pergunta e cada folha sendo uma classificação. O algoritmo que faz a árvore de decisão se baseia num conceito de impureza que pode ser

calculado de duas maneiras gini ou entropia. O algoritmo de *Decision Tree* possui mais hiper-parâmetros, um detalhe é que por padrão ele é configurado para ter profundidade máxima, tendo um *overfit*. Modelos de aprendizado de máquina estão em *overfit* quando não conseguem prever dados novos que nunca tenham passado pelo treinamento isso caracteriza que o modelo decorou os dados.



Fig. 1. Árvore com gini



Fig. 2. Árvore com entropia

## C. SVM

O algoritmo SVM que significa *support vector machine*, é uma algoritmo supervisionado que busca a separação das classes com auxílio do hiperplano que é calculado buscando a melhor divisão entre os pontos de fronteira das classes. Esse calculo pode ser feito de cinco formas diferentes, utilizamos 2 formas a linear que que é uma linha reta entre as fronteiras, a forma rbf que é um acrônimo de *Radial basis function* que pode ser entendida como uma forma arredondada.

## D. Verificação de correlação

O calculo de correlação serve de suporte ao processo de *Feature Enginner*, escolha de características para o aprendizado de máquina.

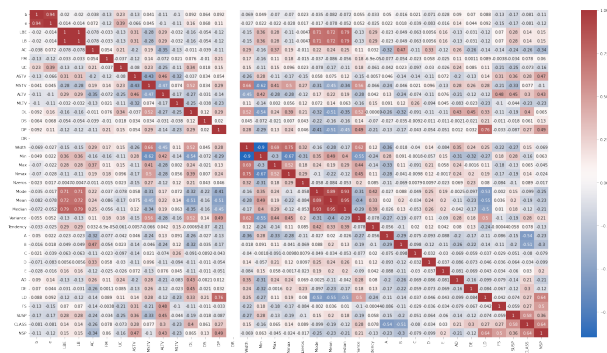


Fig. 3. Correlação do Dataset

algoritmo	métrica	valor
decision_tree gini	Accuracy	98.275862
decision_tree gini	Precision micro:	98.275862
decision_tree gini	Recall micro:	98.275862
decision_tree entropy	Precision micro:	97.962382
SVM linear	Recall micro:	97.962382
SVM linear	Precision micro:	97.962382
decision_tree entropy	Recall micro:	97.962382
SVM linear	Accuracy	97.962382
decision_tree entropy	Accuracy	97.962382
decision_tree entropy	Precision macro:	97.531459
decision_tree gini	Precision macro:	97.486505
decision_tree gini	Recall macro:	97.463634
SVM linear	Precision macro:	97.034949
decision_tree entropy	Recall macro:	96.540739
SVM linear	Recall macro:	96.540739
knn 7	Recall micro:	82.601881
knn 7	Precision micro:	82.601881
knn 7	Accuracy	82.601881
knn 9	Precision micro:	82.288401
knn 9	Accuracy	82.288401
knn 9	Recall micro:	82.288401
knn 5	Accuracy	81.661442
knn 5	Precision micro:	81.661442
knn 5	Recall micro:	81.661442
knn 3	Accuracy	80.564263
knn 3	Precision micro:	80.564263
knn 3	Recall micro:	80.564263
knn 9	Precision macro:	78.027066
SVM rbf	Accuracy	77.742947
SVM rbf	Precision micro:	77.742947
SVM rbf	Recall micro:	77.742947
knn 7	Precision macro:	75.227030
knn 5	Precision macro:	69.635387
knn 3	Precision macro:	62.708588
knn 3	Recall macro:	51.988540
knn 5	Recall macro:	50.924680
knn 7	Recall macro:	49.133176
knn 9	Recall macro:	47.201649
SVM rbf	Recall macro:	33.333333
SVM rbf	Precision macro:	25.914316

## RESULTADOS

Os valores obtidos em cada algoritmo e parâmetro, para cada métrica foram dispostos em tabela e ordenados. Entre os algoritmos a árvore de decisão foi o melhor entre os algoritmos, seguido de SVM e KNN respectivamente.

O algoritmo SVM com o parâmetro rbf teve os piores resultados, enquanto o linear teve resultados satisfatórios acima de 95%, porém não houve uma *tunagem* mais detalhada de parâmetros o apenas a escolha de critério por isso o *underfit*. Não foi solicitado uma melhor *tunagem* dos parâmetros, por esse motivo o algoritmo não teve um bom desempenho. Porém na árvore de decisão como houve uma *tunagem* melhor com a profundidade máxima, pois esse algoritmo é configurado para dar *overfit*, setamos uma profundidade máxima de 10.

