

**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ  
DEPARTAMENTO DE INFORMÁTICA E ELETROTÉCNICA  
ESPECIALIZAÇÃO EM CIÊNCIA DE DADOS E SUAS APLICAÇÕES**

**RODERLEI VENDRAMETTO**

**AVALIAÇÃO DE SINAIS DE SEPSE INFANTO JUVENIL, COM BASE EM  
CIÊNCIA DE DADOS.**

**CURITIBA**

**2021**

**RODERLEI VENDRAMETTO**

**AVALIAÇÃO DE SINAIS DE SEPSE INFANTO JUVENIL, COM BASE EM  
CIÊNCIA DE DADOS.**

**Evaluation of signs of childhood Sepsis, based on Data Science.**

Proposta de Trabalho de Conclusão de Curso apresentado a Especialização em ciência de dados e suas aplicações da Universidade Tecnológica Federal do Paraná, como requisito parcial para a obtenção do título de Especialista.

Orientador: Prof. Dr. Leandro Batista de Almeida.

**CURITIBA**

**2021**



Ministério da Educação  
**UNIVERSIDADE TECNOLÓGICA FEDERAL DO PARANÁ**  
UTFPR - CAMPUS CURITIBA  
DIRETORIA-GERAL - CAMPUS CURITIBA  
DIRETORIA DE PESQUISA E PÓS-GRADUAÇÃO - CAMPUS CURITIBA  
DEPARTAMENTO DE APOIO DAS ESPECIALIZAÇÕES LATO-SENSU DOS  
CURSOS DE INFORMÁTICA - CAMPUS CURITIBA



---

## TERMO DE APROVAÇÃO

### AVALIAÇÃO DE SINAIS DE SEPSE INFATO JUVENIL, COM BASE EM CIÊNCIA DE DADOS

por

**Roderlei Vendrametto**

Este Trabalho de Conclusão de Curso foi apresentado às 20h00min do dia 09 de agosto de 2021 por videoconferência como requisito parcial à obtenção do grau de Especialista em Ciência de Dados e suas Aplicações na Universidade Tecnológica Federal do Paraná - UTFPR - Campus Curitiba. O aluno foi arguido pela Banca de Avaliação abaixo assinados. Após deliberação, a Banca de Avaliação considerou o trabalho aprovado.

---

Prof. Dr. Leandro Batista De Almeida (Presidente)

DAINF - CT - UTFPR

---

Prof. Msc. Christian Carlos De Souza Mendes (Membro)

DAELN - CT - UTFPR

---

Profa. Dra. Rita Cristina Galarraga Berardi (Membro)

DAINF - CT - UTFPR

**O Termo de Aprovação assinado encontra-se no sistema SEI- Processo nº 23064.031212/2021-22**

## **AGRADECIMENTOS**

Se dedicar aos estudos, mesmo depois de um longo tempo afastado, exige um esforço incomum, e nestes momentos, os conselhos de pessoas que querem o seu bem, são fundamentais. Resultado de uma conversa regada a muito incentivo, é que estou concluindo esta especialização, por esta razão, quero deixar aqui todo o meu agradecimento e eterna gratidão, ao meu líder e tutor Laércio Santo de Araújo. Também registro, um agradecimento especial a minha esposa, Débora Zucyszyn Vendrametto, por ter tido e continuar tendo uma paciência e uma dedicação impecável. Torço para que esta monografia sirva de incentivo ao meu filho Ravi Zucyszyn Vendrametto, que enquanto escrevo esta monografia, esta completando um ano de vida.

Deixo aqui minha admiração e agradecimentos ao meu orientador Prof. Dr. Leandro Batista de Almeida, por dispor de seu tempo e sua sabedoria, contribuindo para a condução desta pesquisa.

Não poderia deixar de citar e agradecer, a coordenadora do curso, Prof<sup>a</sup>. Rita Berardi pela sua dedicação e empenho, pelo seu carisma e sorriso cativante.

Enfim, a todos os que por algum motivo contribuíram para a realização desta pesquisa.

“Porquanto somente Eu conheço os planos que determinei a vosso respeito! ”; declara Yahweh, ‘planos de fazê-los prosperar e não de lhes causar dor e prejuízo, planos para dar-vos esperança e um futuro melhor.”  
(Jeremias 29:11).

## RESUMO

Este trabalho apresenta um estudo baseado na tabela de sinais vitais e exames laboratoriais, esta tabela foi definida pelo Instituto Latino Americano da Sepse (ILAS), embasada com centenas de artigos científicos. A partir desta tabela foi criado um algoritmo de alerta, que sinaliza a possibilidade de sepse em pacientes infanto-juvenil, o resultado gerado a partir deste algoritmo, tem baixa acurácia e a proposta é obter de todos os pacientes que em algum momento fizeram parte do alerta, sendo ele positivo ou negativo, colher os sinais estabelecidos pela tabela do ILAS destes pacientes e aplicar ciência de dados, analisando os dados com o algoritmo da árvore de decisão.

**Palavras-chave:** Visualização de dados. Ciência de Dados. Análise de Árvore de decisão. Hospital. Controle de Sepse infantil.

## ABSTRACT

This work presents a study based on the table of visual signs and laboratory tests, this table was defined by the Latin American Institute of Sepsis (ILAS), based on hundreds of scientific articles. From this table, an alert algorithm was created, which signals the possibility of sepsis in children and adolescents. The result generated from this algorithm has low accuracy and the proposal is to obtain from all patients who at some point were part of the alert, whether it is positive or negative, collect the signals identified by the ILAS table of these patients and data science, analyzing the data with the decision tree algorithm.

**Keywords:** Data visualization. Data Science. Decision Tree Analysis. Hospital. Child Sepsis Control.

## LISTA DE ILUSTRAÇÕES

Figura 1 – Fluxograma do Atendimento a suspeita de Sepsis .....	18
Figura 2 – Painel simulando alerta para um paciente com suspeita de Sepsis. ....	19
Figura 3 – Pop-up que é apresentado no sistema PEP para o profissional da assistência.....	20
Figura 4 – Exemplificamos uma árvore de decisão extremamente simples, apenas com dois nós e poucos ramos.....	25
Figura 5 – Tela da ferramenta de extração (kettle) do suíte pentaho vantara, com os steps que foram necessários para a extração do dataset. ....	28
Fotografia 1 – Painel utilizado nas unidades de internação do HOP.....	19
Gráfico 1 - Este gráfico demonstra o percentual de alertas disparados pelo algoritmo, do período de 06/2018 a 06/2021. ....	20
Gráfico 2 – Scatterplot de exames de leucócitos. ....	30
Gráfico 3 – Scatterplot de temperatura. ....	31
Gráfico 4 – Scatterplot de Pressão arterial Sistólica. ....	31
Gráfico 5 – Scatterplot de Frequência Respiratória.....	32
Gráfico 6 – Scatterplot de Frequência Cardíaca.....	32
Gráfico 7 – Árvore de decisão com 60 a 100% de confiança – visão A .....	35
Gráfico 8 – Árvore de decisão com 60 a 100% de confiança – visão B .....	36
Gráfico 9 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão A .....	36
Gráfico 10 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão B .....	37
Gráfico 11 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão C .....	37
Gráfico 12 – Árvore de decisão com 60 a 100% de confiança – visão A .....	38
Gráfico 13 – Árvore de decisão com 60 a 100% de confiança – visão B .....	38
Gráfico 14 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão A .....	39
Gráfico 15 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão B .....	39
Gráfico 16 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão C .....	40
Gráfico 17 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão D .....	40
Gráfico 18 – Árvore de decisão com 60 a 100% de confiança – visão A .....	41
Gráfico 19 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’.....	41
Gráfico 20 – Árvore de decisão com 60 a 100% de confiança – visão A .....	42

Gráfico 21 – Árvore de decisão com 60 a 100% de confiança – visão B .....	42
Gráfico 22 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ .....	43
Gráfico 23 – Árvore de decisão com 60 a 100% de confiança – visão A .....	43
Gráfico 24 – Árvore de decisão com 60 a 100% de confiança – visão B .....	44
Gráfico 25 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ .....	44
Gráfico 26 - Gráfico com o número de infecções por sepse do HOP .....	45

## LISTA DE TABELAS

Tabela 1 - SIRS – Síndrome da Resposta Inflamatória Sistêmica, caracterizada pela presença de dois ou mais dos seguintes sintomas: .....	14
Tabela 2 - SIRS – Síndrome da Resposta Inflamatória Sistêmica, caracterizada pela presença de dois ou mais dos seguintes sintomas: .....	16
Tabela 3 – É demonstrado os dados que serviram como base para a criação da árvore de decisão:.....	25
Tabela 4 – Tabela de totais do datase bruto: .....	29
Tabela 5 – Tabelas de referências de sinais vitais (valores críticos).....	33
Tabela 6 – Tabelas de referências de exame de leucócitos (valores críticos) .....	34
Tabela 7 – Tabela de totais do datas higienizado: .....	34

## **LISTA DE ABREVIATURAS, SIGLAS E ACRÔNIMOS**

### **LISTA DE ABREVIATURAS**

GT	Grupo de Trabalho
HOP	Hospital Objeto da Pesquisa
NQ	Núcleo da Qualidade

### **LISTA DE SIGLAS**

DTI	Departamento de Tecnologia da Informação
UTI	Unidade de Terapia Intensiva

### **LISTA DE ACRÔNIMOS**

ERP	Enterprise Resource Plannig
ETL	Extract, Transform, Load (extração, transformação, carregamento)
ILAS	Instituto Latino Americano da Sepse
PEP	Prontuário Eletrônico do paciente
PL/SQL	Procedural Language/Structured Query Language
SIRS	Síndrome da Resposta Inflamatória Sistêmica

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	<b>13</b>
1.1 O QUE É SEPSE .....	14
1.1.1 Desafios da área de saúde em diagnosticar Sepse infanto-juvenil.....	14
1.1.2 Primeira experiência do HOP com inteligência artificial no diagnóstico de Sepsespe. ....	15
1.1.3. O desafio do Departamento de Tecnologia da Informação do HOP, no desenvolvimento do algoritmo de Diagnóstico de Sepse. ....	16
1.2 . MOTIVAÇÃO. ....	21
1.2.1. Objetivo.....	21
<b>2 CONCEITO E FERRAMENTAS</b> .....	<b>21</b>
2.1 O QUE É CIÊNCIA DE DADOS? .....	22
2.1.1 Algoritmos de Machine Learning.....	22
2.1.1.1 Algoritmos supervisionados .....	22
2.1.1.2 Algoritmos não supervisionados .....	23
2.1.1.3 Algoritmos de Arvore de Decisão.....	24
2.2 FERRAMENTAS .....	26
2.2.1 Ferramenta de Data Integration – Kettle da suíte pentaho vantara .....	26
2.2.2 BIGML.....	26
<b>3 DESENVOLVIMENTO</b> .....	<b>27</b>
3.1 COLETA DE DADOS .....	27
3.1.1 Técnica para extração dos dados .....	27
3.1.2 Análise dos dados brutos.....	29
3.1.3 Processo de Higienização. ....	33
3.2 RESULTADOS.....	35
3.2.1 Resultado da análise de pacientes na faixa etária de 28 a 90 dias. ....	35
3.2.2 Resultado da análise de pacientes na faixa etária de 90 a 365 dias de vida. ....	38
3.2.3 Resultado da análise de pacientes na faixa etária de 366 a 1825 dias. ....	41
3.2.4 Resultado da análise de pacientes na faixa etária de 1825 a 4380 dias. ....	42
3.2.5 Resultado da análise de pacientes na faixa etária de 4380 a 6570 dias. ....	43
3.3 DADOS DO HOSPITAL OBJETO DA PESQUISA (HOP).....	44
<b>4 CONCLUSÃO (OU CONSIDERAÇÕES FINAIS)</b> .....	<b>46</b>
<b>REFERÊNCIAS</b> .....	<b>47</b>

## 1 INTRODUÇÃO

Machine learning faz parte da Engenharia e da ciência da computação que evoluiu do estudo de reconhecimento de padrões e da teoria do aprendizado computacional em inteligência artificial. Estes algoritmos trabalham construindo modelos a partir de massa de dados amostrais a fim de fazer previsões ou decisões guiadas pelos dados ao invés de simplesmente seguindo estáticas e inflexíveis instruções programadas.

Com base no conhecimento obtido nas aulas da especialização, percebemos que a aplicação dos algoritmos de Machine learning, poderia contribuir para o dia a dia dos profissionais de saúde, através de ferramentas de ajuda à decisão.

Foi escolhido por validar um algoritmo existente, de instruções programadas no método inflexível, que avalia as condições do paciente e com base em parâmetros, é disparado um alerta sinalizando os profissionais de saúde, a possibilidade de SEPSE.

A comunidade da saúde tem inúmeros desafios, e uma delas é o combate a sepse, a sepse pode surgir por uma infecção, o combate exige rápido diagnóstico para que seja possível dar início ao tratamento, uma vez diagnosticado com sepse, o paciente sem o devido tratamento pode vir a óbito em questão de horas. Em busca de acelerar este diagnóstico, um sistema de alertas foi desenvolvido pela equipe do departamento de tecnologia da informação (DTI) do hospital objeto da pesquisa (HOP), hospital que esta localizada na cidade de Curitiba, no estado do Paraná. Este sistema vem servindo de ferramenta ao corpo clínico e de enfermagem desde 2018.

A linha de pesquisa escolhida, esta embasada nos critérios de diagnóstico aplicada a este sistema, tendo como público-alvo, pacientes de 0 a 18 anos de idade, com o objetivo de reavaliar os sensores, se utilizando algoritmos de ciência de dados.

## 1.1 O QUE É SEPSE

Com base no instituto Latino Americano de Sepse (ILAS), sepse é a Resposta Inflamatória Sistêmica (SIRS), pode ser definida pela presença de uma disfunção orgânica que ameaça à vida, resposta do corpo humano desregulado pelo hospedeiro a uma infecção, seja causada por vírus, fungos, bactérias ou protozoários e caracterizada por manifestações múltiplas (ILAS, 2015).

Sepse, também é conhecida como sépsis, sepsia, que é entendida como uma infecção grave pode afetar um órgão do corpo humano ou todo o sistema, era conhecida antigamente como septicemia ou infecção no sangue, hoje é mais conhecida como infecção generalizada.

A cada hora de atraso na administração do antibiótico aumenta o risco de mortalidade em cerca de 8%. (Manual de Abordagem Sepse, MS, Albert Einstein, ILAS).

O diagnóstico de sepse é conhecido como disfunção ou falência de múltiplos órgãos. É responsável por 25% da ocupação de leitos em Unidades de Terapia Intensiva (UTI) no Brasil. Atualmente a sepse é a principal causa de morte nas UTIs e uma das principais causas de mortalidade hospitalar tardia. Tem alta mortalidade no país, chegando a 65% dos casos, enquanto a média mundial está em torno de 30-40%. Disponível em: <https://ilas.org.br/o-que-e-sepse.php>. Acesso em: 8 de maio de 2021.

### 1.1.1 Desafios da área de saúde em diagnosticar Sepse infanto-juvenil.

Embora o ILAS disponibilize ampla bibliografia tratando da sepse, grande parte desta documentação se destina ao diagnóstico e tratamento exclusivamente de adultos, por esta razão, o diagnóstico da sepse em adultos se torna menos complexa, veja na tabela 1, os critérios de diagnóstico:

**Tabela 1 - SIRS – Síndrome da Resposta Inflamatória Sistêmica, caracterizada pela presença de dois ou mais dos seguintes sintomas:**

Sintomas	Resultado
Temperatura Central	> 38,3° C ou < 36° C ou equivalente em termos de temperatura axilar.
Frequência Cardíaca	> 90 bpm.
Frequência Respiratória	> 20 rpm, ou PaCO <sub>2</sub> < 32mmhg

Leucócitos totais > 12.000/mm<sup>3</sup>; ou < 4.000/mm<sup>3</sup> ou presença de > 10% de formas jovens (desvio à esquerda).

---

**Fonte: Protocolo de tratamento – ILAS – Adulto (2018)**

Porém quando é tratado de crianças e adolescentes, estes parâmetros contribuem, mas não é suficiente para um diagnóstico preciso, a tabela sofre variações com base na idade do paciente, e mesmo assim, ainda hoje existem poucos artigos e/ou bibliografias tratando de sepse infanto-juvenil.

#### 1.1.2 Primeira experiência do HOP com inteligência artificial no diagnóstico de Sepsis.

O HOP, esta sempre em busca da inovação tecnológica, onde procura instrumentar seu corpo clínico e de enfermagem, com tecnologia de ponta para obter o melhor diagnóstico e assim melhorar a qualidade no tratamento de seus pacientes.

Alinhado com o objetivo da inovação, o HOP aceitou a proposta de uma empresa local, que oferecia gratuitamente a instalação de um robô com inteligência artificial para diagnosticar a sepse, esta empresa divulgou ter um case de sucesso em um hospital da região.

Após benchmarking, no ano de 2017 foi preparado o ambiente sistêmico, plugando o robô ao Enterprise Resource Planning (ERP) Hospitalar, sistema que recebe os registros das equipes multidisciplinares na assistência à saúde, como por exemplo: medicação, sinais vitais e resultados de exame.

O hospital trabalhou amplamente na divulgação do projeto internamente, onde preparou toda a instituição para a chegada do robô com inteligência artificial, despendeu várias horas treinando equipes para analisar e entender os sinais que seriam disponibilizados pelo robô, e por tanto, a instituição estava preparada.

O projeto percorreu um pouco mais de um ano, com exaustivas reuniões e reajustes de sensores do robô, porém ao perceberem que os alertas do robô apresentavam muitos falsos alertas do que positivos, acabou minando a confiança da equipe de corpo clínico e de enfermagem que solicitaram a desativação do robô.

### 1.1.3. O desafio do Departamento de Tecnologia da Informação do HOP, no desenvolvimento do algoritmo de Diagnóstico de Sepsis.

A instituição vinha de ampla divulgação e de inúmeros treinamentos para interpretação dos sinais de sepsis emitido pelo robô, todo o corpo clínico estava preparado, os colaboradores ansiosos para a chegada do “robô com inteligência artificial”, porém os inúmeros alertas falsos emitidos pelo robô gerou desconfiança levando o projeto a ser interrompido. Sem o projeto e com a instituição com o anseio por tecnologia, a diretoria então reuniu o setor de Núcleo da Qualidade (NQ) e a DTI, apresentando o desafio de desenvolver o seu próprio robô de diagnóstico de sepsis.

Com o desafio aceito, uma parceria foi selada entre representantes do corpo clínico especialistas em sepsis, NQ e DTI, sendo criado o Grupo de Trabalho de sepsis.

O GT de sepsis, utilizando como base artigos e várias reuniões, estabeleceram com base na tabela do ILAS 2016 e na experiência na assistência dos profissionais médicos do HOP, ajustaram a tabela do ILAS para o mais próximo da realidade e da lógica assistencial infantil.

**Tabela 2 - SIRS – Síndrome da Resposta Inflamatória Sistêmica, caracterizada pela presença de dois ou mais dos seguintes sintomas:**

Faixa Etária	Taquicardia – batimentos por minute (bpm)	Bradycardia	Frequência Respiratória (cpm)	Contagem de leucócitos (10 <sup>3</sup> /mm <sup>3</sup> )	Pressão Arterial Sistólica (mmHg)	Hipotermia Febre
> 0 a 7 dias	> 190	< 100	> 70	> 34.000	< 60	
> 7 dias a 28 dias	> 190	< 100	> 70	< 5.000 ou > 19.500	< 60	< 35,0° ou >= 38,0°
> 28 a 90 dias	> 190	< 100	> 70	< 5.000 ou > 19.500	< 70	
> 90 a 365 dias	> 190	< 90	> 60	< 5.000 ou > 17.500	< 70	
> 365 a 1825 dias	> 140	NA	> 34	< 6.000 ou > 15.500	< 70	
> 1825 a 4380 dias	> 130	NA	> 30	< 4.500 ou > 13.500	< 83	
> 4380 a 6570 dias	> 110	NA	> 16	< 4.500 ou > 11.000	< 90	

Fonte: Protocolo de tratamento – ILAS – Infanto-juvenil (2016) – Modificada

A tabela 2, só compreende pacientes de 0 dias de vida, até 18 anos ou 6570 dias, que é o público alvo da pesquisa.

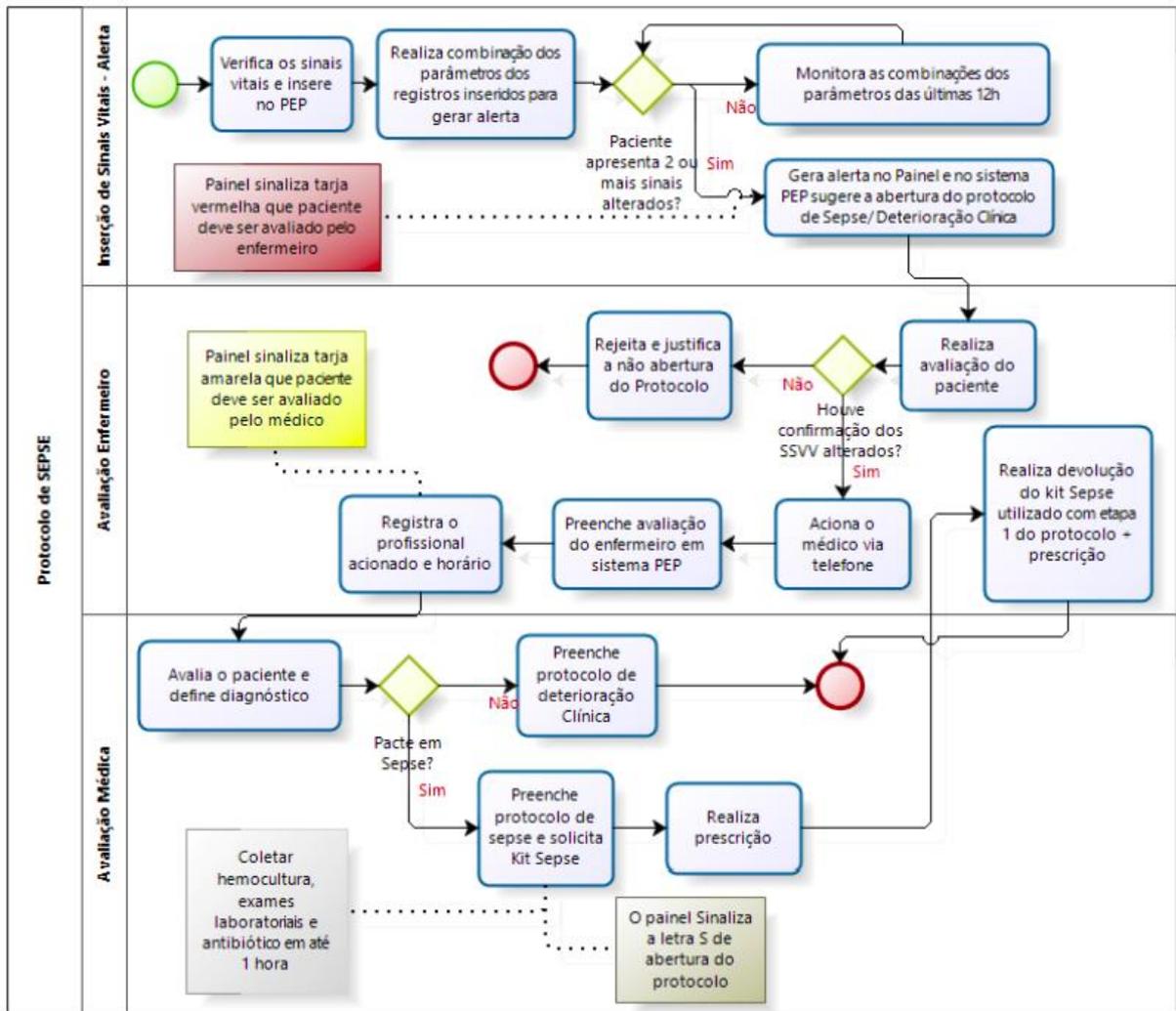
Na ocasião que foi apresentado a tabela 2, como o DTI não disponha de nenhum membro com conhecimento técnico em ciência de dados, decidiram por desenvolver o algoritmo com ferramentas de seu domínio, foi escolhido utilizar linguagem Procedural Language/Structured Query Language (PL/SQL).

Foi criada uma sequência de algoritmo (funções), que leva em consideração a idade do paciente, pois para cada idade existe um parâmetro diferente, com base neste parâmetro são avaliados os registros clínicos das últimas 12 horas, como resultado de exames de sangue de leucócitos, batimento cardíaco, frequência respiratória e pressão arterial, avaliando se estão alterados com base no parâmetro, caso apresentem 2 sinais alterados, o algoritmo retorna um alerta.

Em cada unidade de internação ou pronto atendimento o HOP tem uma televisão com um minicomputador conectado a ela, um navegador de internet está aberto com uma URL do site do painel, que também foi desenvolvido pelo DTI, este painel serve como recurso para o corpo clínico e para a enfermagem, é apresentado nele os pacientes internados naquela unidade, e apresenta informações complementares como, leito, idade, exames solicitados/realizados ou se está de alta médica, entre outras informações. Neste painel, quando o algoritmo alerta dois sinais alterados, a cor da linha do paciente é alterada para vermelha.

No sistema ERP Hospitalar, módulo de assistência ao paciente, chamado de prontuário eletrônico do paciente (PEP), quando o paciente é alertado aparece um pop-up, onde os enfermeiros devem aceitar ou descartar o alerta com base na avaliação do paciente. Todo o fluxo do protocolo de sepse pode ser visto no fluxograma apresentado na figura 1.

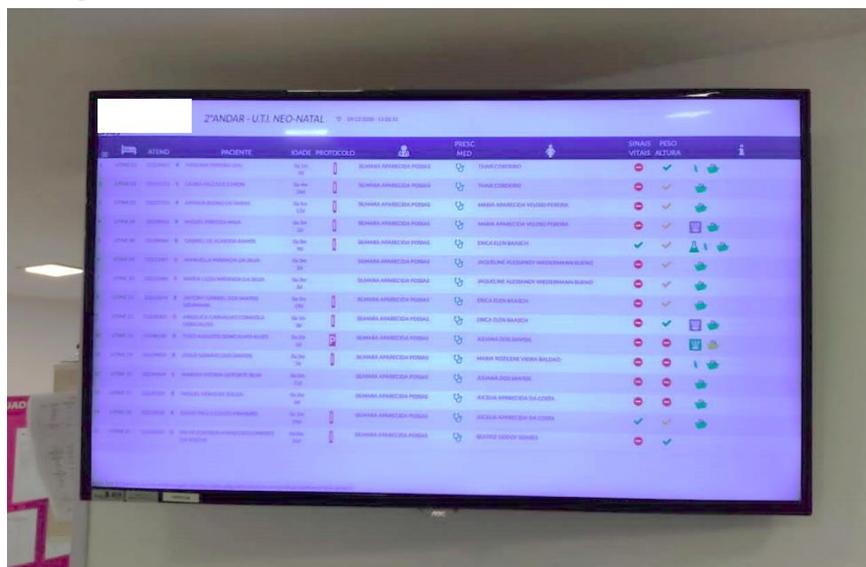
Figura 1 – Fluxograma do Atendimento a suspeita de Sepses



Fonte: Protocolo de Sepses do HOP (2015)

Na fotografia 1, é possível visualizar o painel da unidade.

**Fotografia 1 – Painel utilizado nas unidades de internação do HOP.**



Fonte: Autoria própria

Na figura 2, podemos visualizar um painel simulando um alerta, o nome do paciente é pintado de vermelho;

**Figura 2 – Painel simulando alerta para um paciente com suspeita de Sepsis.**

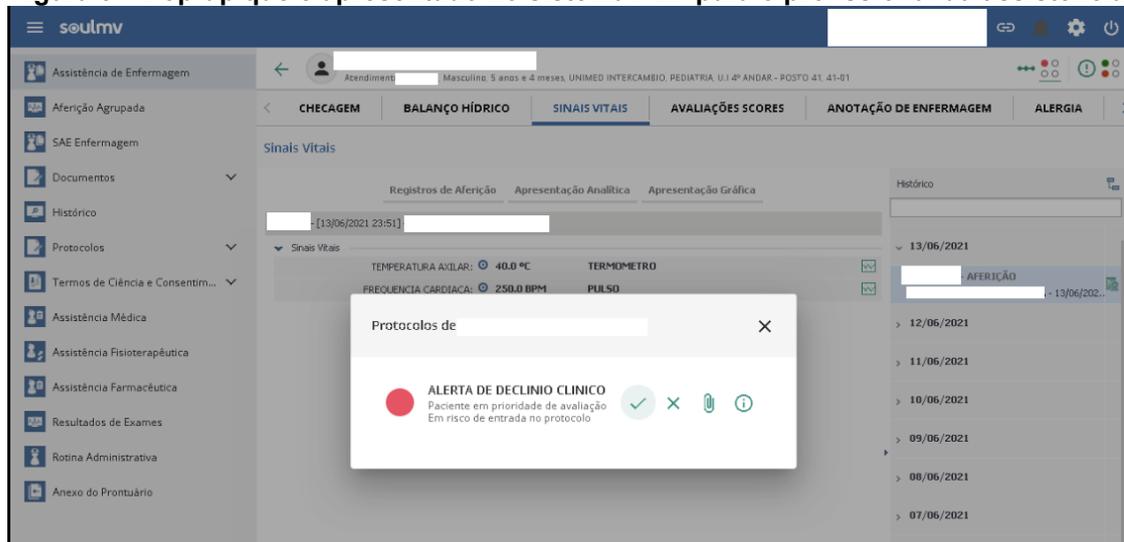
1ºANDAR -POSTO 11

	ATEND	PACIENTE	IDADE	PROTOCOLO	PRESC MED	SINAIS VITAIS	PESO ALTURA
1	101-01 000000	NOME DO PACIENTE	9a 8m 24d	I + P	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
2	102-01 000000	NOME DO PACIENTE	11a 3m 1d	P	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ -
3	102-02 000000	NOME DO PACIENTE	2a 9m 10d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
4	103-01 000000	NOME DO PACIENTE	6a 0m 23d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
5	103-02	NOME DO PACIENTE	16a 5m 2d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	- -
6	104-01 000000	NOME DO PACIENTE	12a 9m 20d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
7	104-02 000000	NOME DO PACIENTE	8a 4m 3d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
8	104-03 000000	NOME DO PACIENTE	13a 5m 22d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
9	105-01 000000	NOME DO PACIENTE	14a 3m 10d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓
10	105-02 000000	NOME DO PACIENTE	17a 1m 6d	I	NOME DO MÉDICO RESPONSÁVEL	NOME DO TÉC. DE ENFERMAGEM RESPONSÁVEL	✓ ✓

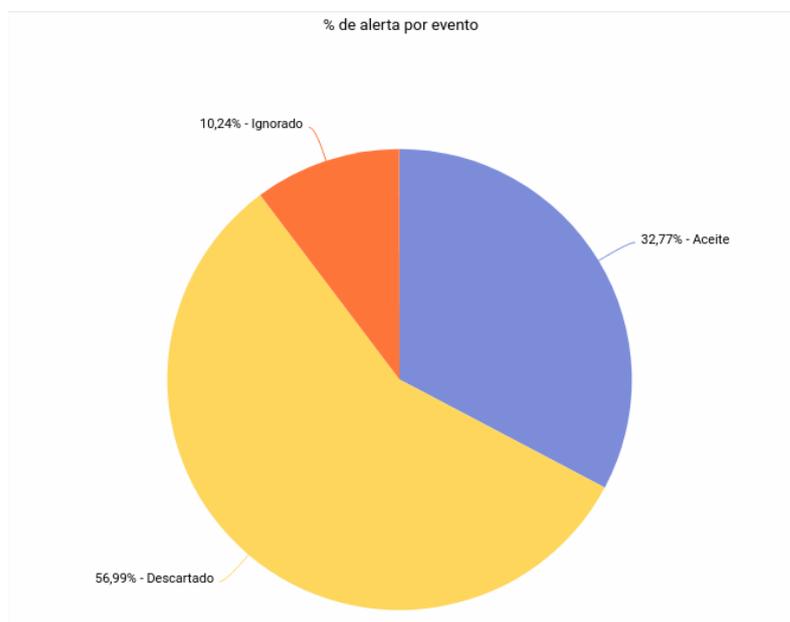
Fonte: Autoria própria

Na figura 3 é possível visualizar o pop-up do alerta que aparece para o profissional da assistência;

**Figura 3 – Pop-up que é apresentado no sistema PEP para o profissional da assistência.**



Fonte: Autoria própria



**Gráfico 1 - Este gráfico demonstra o percentual de alertas disparados pelo algoritmo, do período de 06/2018 a 06/2021.**

Fonte: Departamento do Núcleo da Qualidade do (HOP)

Legenda:

Ignorado → O sinal foi ignorado pelo corpo clínico / enfermagem.

Descartado → O paciente foi avaliado, porém não se trata de sepse.

Aceite → O paciente foi submetido ao protocolo de sepse.

O gráfico demonstra uma taxa de acurácia relativamente baixa, porém mesmo com o acerto de somente 32,77% dos casos, tanto o corpo clínico, como a enfermagem estão otimistas, pois antes deste mecanismo, era preciso confiar no tato da enfermagem, que infelizmente nem sempre conseguiam identificar sepse no tempo certo.

## 1.2 . MOTIVAÇÃO.

A assistência à saúde é complexa por natureza, a saúde é uma das áreas que mais recebem investimento para pesquisas, porém mesmo com tantos investimentos ainda carece de tecnológica. Para alguns hospitais públicos, a tecnologia pode até existir no mercado, mas acaba sendo inviável financeiramente. A eficiência no atendimento ao paciente exige profissionais dedicados e experientes, mas isso não está sendo mais o suficiente, tecnologia cada vez mais faz parte do dia a dia do Profissional de saúde.

O desafio da área de saúde em lidar com tantos vírus, drogas, procedimentos, os obriga a ter um apoio à decisão, atualmente este apoio é feito por telemedicina, videoconferência, telefone e até mesmo por whatsapp, mas nem sempre tem outra pessoa do outro lado para dividir o problema, e mesmo este grupo de pessoas corre o risco de não acertar em 100% a melhor conduta, sendo assim abrindo possibilidade para aplicação de ciência de dados.

### 1.2.1. Objetivo.

O objetivo deste trabalho é coletar os sinais vitais do paciente como os resultados de exames, no período de 12 horas, anterior a data dos alertas gerado pelo algoritmo comum de avaliação existente. Com base nesta massa de dados, e pacientes que deram positivo e negativo para Sepse, aplicar algoritmos de Machine learning e analisar os resultados.

## 2 CONCEITO E FERRAMENTAS

Neste capítulo, abordaremos temas e aspectos relevantes em Ciência de Dados, esta abordagem estará tratando de conceitos, definições e terminologias. Além da ferramenta escolhida para análise.

## 2.1 O QUE É CIÊNCIA DE DADOS?

Ciência de dados em inglês: data Science, trata dados estruturados e não estruturados, é uma extensão da evolução da estatística e os estudos são voltados na sua grande maioria para a análise de grandes massas de dados (Data lake), buscando definir o perfil social, econômico e financeiro, sempre procurando obter conhecimento, gerando padrões para possíveis tomadas de decisão.

As principais coisas que definem um cientista de dados além de um estatístico, a capacidade de trabalhar com Big Data e experiência em aprendizado de máquina, computação e construção de algoritmos. Digamos que um cientista de dados seja alguém que extrai conhecimento de dados desorganizados. O mundo esta cheio de pessoas tentando transformar dados em conhecimento. (GRUS, 2016)

### 2.1.1 Algoritmos de Machine Learning

Machine learning (Aprendizado de máquina), é possibilitar que a máquina tenha a habilidade de aprender, mesmo que nem tudo tenha sido programado (ensinado) previamente.

Destes algoritmos temos duas divisões, uma sendo aprendizado de máquina supervisionada e outra não supervisionada.

#### 2.1.1.1 Algoritmos supervisionados

Os algoritmos de aprendizagem supervisionada são baseados em dados rotulados. É apresentada a máquina uma carga de dados rotulados para treinamento e o resultado obtido é com base nos dados aprendido no treinamento.

Aprendizagem de máquina supervisionada ainda pode ser dividida em mais dois grupos:

- **Classificação:** São os casos conhecidos e previsíveis, como por exemplo, um surfista gosta de surfar quando há sol, logo sabemos que é preciso ter vento para ter ondas e por sua vez, se o algoritmo entende que quando tiver sol, mas não estiver ventando, naturalmente não terá ondas e por sua vez o surfista não conseguirá surfar. Quando não tiver sol, mas se estiver ventando, o surfista não pegará onda, pois ele gosta de pegar ondas somente quando há sol. Isso é chamado de algoritmos de classificação.
- **Regressão:** Existem vários algoritmos de regressão, como regressão linear ou regressão logística, a mais utilizada é a regressão linear, normalmente é para prever algo, como por exemplo, a temperatura, com base em rótulos (histórico), é possível prever a temperatura da próxima semana, mas também é possível envolver outras variáveis que pode influenciar na previsibilidade.

Neste grupo, também se encaixa os algoritmos de árvore de decisão, trataremos deste tema mais a frente.

#### 2.1.1.2 Algoritmos não supervisionados

Nem sempre temos dados rotulados, como por exemplo: em um mercado, podemos ter consumidores que gostam de comprar cerveja e queijo, já outros gostam de comprar pipoca e refrigerante, se por ventura você deixar mais distante um produto do outro, você pode acabar tendo um aumento das vendas, mas neste caso, somente algoritmos não supervisionados conseguem definir este perfil de comprador. Pois neste caso, o algoritmo não terá dados de aprendizado, ele terá que definir o perfil de cada comprador e traçar um padrão.

Aprendizagem de máquina supervisionada ainda pode ser dividida em mais dois grupos:

- **Algoritmos de agrupamento:** São também chamados de clustering, normalmente agrupam valores, com base em informações preestabelecidas, estabelecendo padrões.

- Algoritmos de transformação: normalmente é utilizado para melhorar o desempenho de outros algoritmos de aprendizagem de máquina, ou simplesmente para facilitar a interpretação humana.

### 2.1.1.3 Algoritmos de Arvore de Decisão

Para o aprendizado de máquina existem diversos métodos, e neste trabalho estamos abordando o aprendizado supervisionado para a indução de arvores de decisão. Pode se considerar uma árvore de decisão, como uma estrutura hierárquica contendo nós e links. Uma análise de decisão poderia ser tomada com base em uma tabela, porém a forma alternativa de expressar as regras da tabela é mais bem interpretada quando esta no formato de árvore.

Trabalhar com árvore de decisão, atualmente é um dos métodos mais utilizado para resolução de problemas de classificação, tomada de decisão e aprendizagem de máquina (MITCHELL, 1997).

Uma arvore de decisão é composta pelo 1º nó, também chamado como raiz, a partir deste nó temos as ramificações, que podemos denominar links ou ramos, que segue para o(s) próximo(s) nós, os últimos nós são chamados de terminais ou folhas. A cada nó tem um valor, um resultado de teste.

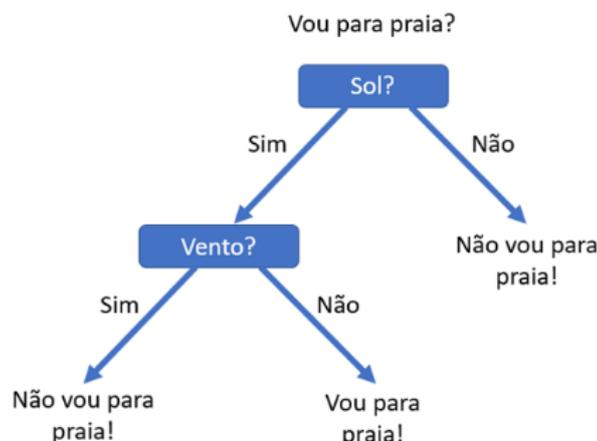
Um dos métodos muito utilizado é o Top Down Induction Decision Tree (TDIDT) este método apresenta um processo recursivo para a construção de árvores, com base em um conjunto de exemplos de treinamento em um nó, apresenta o nó como folha, ou divide o conjunto em um subconjunto.

A divisão de conjuntos é uma das características que diferencia os algoritmos de indução de árvores. TDIDT é base para vários outros algoritmos de arvores de decisão, os mais populares são C4.5, ID3 e CART.

O propósito destes algoritmos é apresentar no primeiro nó o melhor atributo, para em seguida distribuir os principais conjuntos. Segundo Mitchell(1997) para obter árvores enxutas com dados mais consistentes (ou como Mitchell diz, “puro”), são necessário ter dados mais precisos ou concisos para o teste.

Na figura 4, é demonstrado com um exemplo simples como funciona uma árvore de decisão:

**Figura 4 – Exemplificamos uma árvore de decisão extremamente simples, apenas com dois nós e poucos ramos.**



Fonte: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/> Acesso em 28/06/2021

A figura 4 ilustra uma árvore de decisão, a sua interpretação é muito mais fácil compreensão do que a tabela apresentada na tabela 3:

**Tabela 3 – É demonstrado os dados que serviram como base para a criação da árvore de decisão:**

Dia	Sol?	Vento?	Vou para praia?
1	Sim	Sim	Não
2	Sim	Sim	Não
3	Sim	Não	Sim
4	Não	Não	Não
5	Não	Sim	Não
6	Não	Sim	Não

Fonte: <https://didatica.tech/como-funciona-o-algoritmo-arvore-de-decisao/> Acesso em 28/06/2021

A simplicidade desta ilustração é justamente para tentar passar com clareza o funcionamento de uma árvore de decisão, se analisarmos a árvore de decisão, percebemos que são necessário dois fatores positivos para a pessoa ir para a praia, é preciso ter sol e estar ventando, caso contrário ela não ira. Note que ler estas informações na tabela requerer um pouco mais de dedicação para chegar a este critério, onde na árvore é apresentado com maior clareza.

Naturalmente, uma análise desta nem seria necessário criar ou aplicar um algoritmo, porém quando estamos tratando de centenas e de dados de qualidade, a árvore de decisão tem o mesmo desempenho.

## 2.2 FERRAMENTAS

Hoje em dia existem inúmeras ferramentas para ciência de dados, cada uma com sua especificidade, voltada para um determinado objetivo. Para este trabalho foi utilizado para extract – transform – load (ETL) o kettle da suíte pentaho da empresa Hitachi Vantara. E para aplicação de algoritmo de data Science, escolhemos a utilização da ferramenta BIGML / WEKA.

### 2.2.1 Ferramenta de Data Integration – Kettle da suíte pentaho vantara

A Pentaho suíte oferece um conjunto de ferramentas, para a fácil integração e movimentação de dados, trata-se de uma plataforma moderna, que facilita o uso pela TI e até mesmo para os usuários finais, fazer análise do negócio utilizando ferramentas de BI. Na Pentaho suíte, é possível criar visualizações e explorar facilmente todos os dados que impactam os resultados do negócio. Pentaho Kettle permite a TI e desenvolvedores acessarem, higienizar e integrar dados de qualquer fonte, e os entregarem aos seus aplicativos de negócios, tudo de dentro de uma ferramenta gráfica, intuitiva e fácil de uso.

### 2.2.2 BIGML

O software BigML foi fundado em 2011, a missão foi tornar o aprendizado de máquina fácil para todos, onde qualquer pessoal seria capaz de construir soluções sofisticadas baseadas em aprendizado de máquina de maneira acessível, utilizando padrões preditivos de seus dados ou de qualquer pessoa.

Em um futuro próximo, todos os aplicativos serão preditivos. Para permanecerem relevantes em um mundo mais rápido, complexo, conectado e incerto, os aplicativos preditivos terão que aproveitar as vantagens do Aprendizado de Máquina e outras técnicas de Inteligência Artificial de uma forma ou de outra. A plataforma, as implantações privadas e o rico conjunto de ferramentas do BigML continuarão a ajudar a criar, experimentar rapidamente, automatizar totalmente e gerenciar fluxos de trabalho de aprendizado de máquina que impulsionam os

melhores aplicativos inteligentes da classe. Disponível em: <https://bigml.com/about/>. Acesso em: 14 de junho de 2021.

Foi escolhido utilizar a ferramenta, justamente pelo propósito que ela foi criada, pelo recurso de modelagem automatizada ou (Auto ML), com baixa codificação (low code), é possível obter resultados confiáveis, que vem de encontro com o curto espaço de tempo de execução deste trabalho de conclusão de curso.

### **3 DESENVOLVIMENTO**

A linha de pesquisa se baseia na tabela de parâmetros disponibilizados pelo ILAS, parâmetros estes que sofreram leve modificação pelo GT de sepse do HOP, esta tabela apresenta parâmetros diferentes conforme a faixa etária do paciente.

#### **3.1 COLETA DE DADOS**

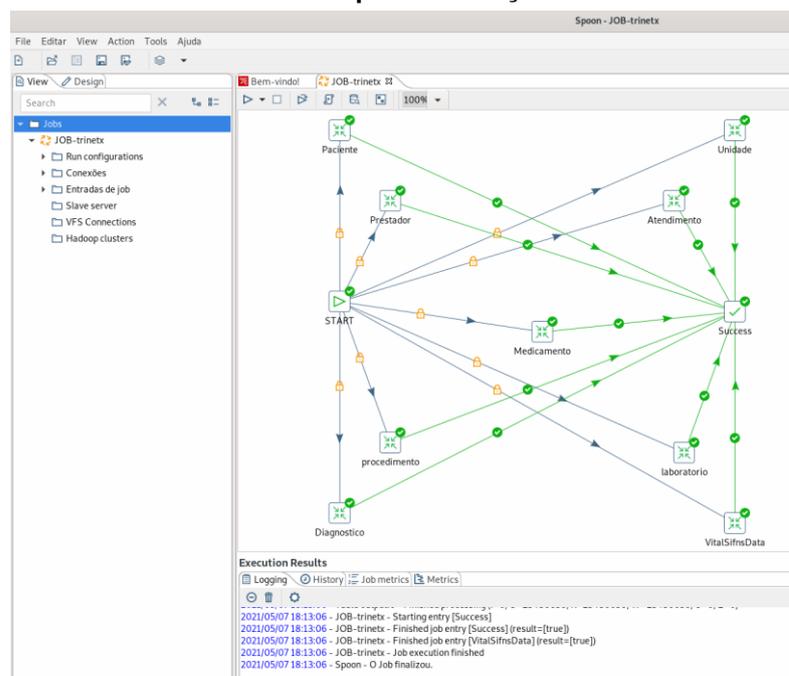
Existem várias maneiras de registrar as informações no sistema, para o registro do dado cadastral do paciente e o registro de atendimento (Por exemplo: o ato de internar) são imputados no sistema pelo administrativo, já os sinais vitais do paciente são imputados pela enfermagem. Já os resultados laboratoriais existem rotinas automáticas de integração, o sistema ERP Hospitalar comunica de maneira bidirecional, encaminhando o pedido do exame para o equipamento e tendo como o retorno o resultado do exame do paciente após liberação por um bioquímico. E por último, todo alerta gerado no sistema, através do algoritmo criado pela DTI, é armazenado no banco de dados. Todos os dados estão reunidos e armazenados no banco de dados relacional do sistema ERP Hospitalar.

##### **3.1.1 Técnica para extração dos dados**

O Data set é uma coleção de dados, sendo que um paciente pode ter tido mais que um alerta em seu atendimento, por tanto, um paciente pode ter mais uma linha no dataset, nesta linha iremos compor com resultados de exame de sangue de leucócitos, e aferições das últimas 12 horas, levando em consideração o horário de registro do alerta de sepse.

Pela complexidade em como os dados está armazenado no banco de dados, um conjunto de SQL com união de tabelas não satisfazem, por esta razão foi utilizado o sistema data integration (Kettle) do suíte da pentaho hitachi vantara.

**Figura 5 – Tela da ferramenta de extração (kettle) do suíte pentaho vantara, com os steps que foram necessários para a extração do dataset.**



Fonte: Autoria própria

Resultado da transformação foi gerado um arquivo com formato csv, que será utilizado como fonte de dados para análise. Demonstramos um trecho do arquivo CSV;

```
= "sepse", "faixaetaria", "ex_hemograma_leucocitos_min", "ex_hemograma_leucocitos_max", "pressao_sistolica_min", "frequencia_respiratoria_max", "frequencia_cardiaca_min", "frequencia_cardiaca_max", "temperautra_min", "temperautra_max"
= "N", "366a1825dias", "", "", "64", "48", "110", "171", "35,7", "38"
= "N", "91a365dias", "", "", "60", "46", "95", "172", "36", "37"
= "N", "4381a36500dias", "", "", "80", "35", "75", "115", "35,2", "37,8"
= "N", "366a1825dias", "", "", "81", "36", "100", "145", "35,1", "37"
```

### 3.1.2 Análise dos dados brutos

Na tabela 4, é apresentada os totais de registros, nesta visão, os dados na condição que vieram do banco de dados relacional do sistema ERP Hospitalar.

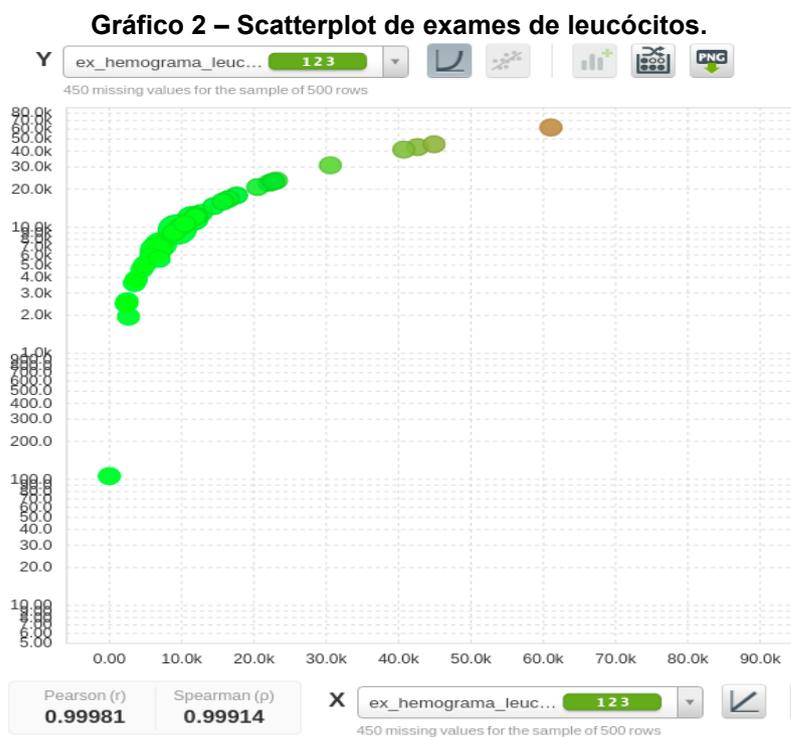
**Tabela 4 – Tabela de totais do datase bruto:**

Faixa Etária	Pacientes com sinais de Sepsis?	
	Não	Sim
> 0 a 7 dias	1	4
> 7 dias a 28 dias	21	15
> 28 a 90 dias	184	32
> 90 a 365 dias	441	81
> 365 a 1825 dias	2419	209
> 1825 a 4380 dias	1552	170
> 4380 a 6570 dias	3403	113
<b>Total</b>	<b>8021</b>	<b>624</b>

**Fonte: Autoria Própria**

Este data set foi carregado na aplicação do BIGML, e com o recurso gráfico de scatterplot, analisamos as possíveis sujeiras nos dados, entendido daqui em diante como outliers.

No gráfico 2, é possível identificar os outliers pelos pontos mais afastados da curva central.



Fonte: Autoria Própria

No gráfico de temperatura, conseguimos observar com clareza os pontos concentrados na faixa de valores esperado, e alguns pontos mais distantes, no gráfico localizado no canto inferior do gráfico, entendemos como outliers.

Gráfico 3 – Scatterplot de temperatura.



Fonte: Autoria Própria

A preocupação da assistência com o dado de pressão arterial sistólica é sempre de estar baixa a cima do permitido. Onde sabemos que o limite seria de 25 mmHg, no gráfico 4, demonstra alguns pontos que são inferior aos 25 mmHg, estes valores são entendidos como outliers.

Gráfico 4 – Scatterplot de Pressão arterial Sistólica.



Fonte: Autoria Própria

A apresentação do gráfico de frequência respiratória tem uma demonstração clara do conjunto de dados esperado com os seus valores de outliers, note que no gráfico é possível visualizar um grande círculo que estabelece os valores dentro do

padrão, sendo que tem um ponto um pouco mais distante, este ponto é claramente um registro de outlier.

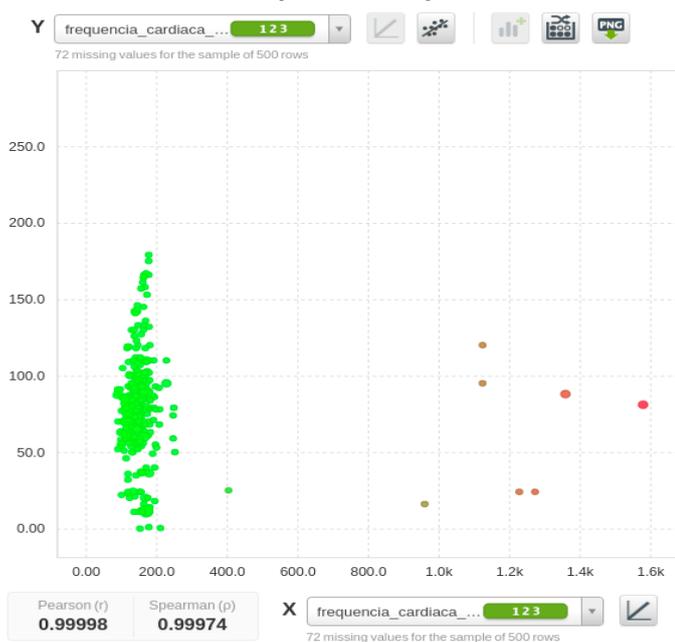
**Gráfico 5 – Scatterplot de Frequência Respiratória.**



Fonte: Autoria Própria

O gráfico 6, demonstra uma aglomeração de pontos no canto inferior esquerdo, os pontos não verdes, que estão posicionados mais ao canto direito, também são tratados como outliers.

**Gráfico 6 – Scatterplot de Frequência Cardíaca.**



Fonte: Autoria Própria

Embora aparentemente sejam poucos dados fora do padrão, se recomenda a limpeza dos dados, para uma melhor precisão no resultado, por isto, estaremos tratando de higienização no próximo item. De acordo com (FAYYAD, 1996), na etapa de higienização as operações básicas incluem, a remoção de ruídos e desvios (se possível), a coleta das informações necessária para o modelo ou a adequação do ruído, traçando e definindo as estratégias para lidar com campos incompletos, ou fora do padrão.

### 3.1.3 Processo de Higienização.

Analisando os dados, é possível identificar registros totalmente fora do esperado, por exemplo: temperatura de 300 graus, uma temperatura corporal desta para um ser vivo é impossível, pois a ebulição inicia com 100 graus. Embora a maioria dos dados de outliers seja possível identifica-lo e perceber que houve erro de digitação, buscamos referências bibliográficas para amparar a higienização do Data set.

Os valores apresentados na tabela 5 entende-se como o limite aceito de valores em suas extremidades, tendo ciência que um paciente crítico, pode ultrapassar este valor, decidiu por aplicar a higienização no maior e menor valor informado, acrescido de 10%, desta forma, garantindo que estaremos tratando também pacientes em condições críticas.

**Tabela 5 – Tabelas de referências de sinais vitais (valores críticos)**

Faixa Etária	Frequência Cardíaca		Frequência Respiratória	Pressão Arterial	Temperatura
	Em vigília	Em sono			
> 0 a 7 dias	100 a 205 bpm	90 a 160 bpm	30 a 53 rpm	< 60 mmHg	< 35° ou > 41°
> 7 dias a 28 dias	100 a 180 bpm	90 a 160 bpm	30 a 53 rpm	< 60 mmHg	
> 28 a 90 dias	98 a 140 bpm	80 a 120 bpm	30 a 53 rpm	< 70 mmHg	
> 90 a 365 dias	80 a 120 bpm	65 a 100 bpm	22 a 37 rpm	< 70 mmHg	
> 365 a 1825 dias	80 a 120 bpm	58 a 90 bpm	20 a 28 rpm	< 70 mmHg	
> 1825 a 4380 dias	75 a 118 bpm	58 a 90 bpm	18 a 25 rpm	< 83 mmHg	

> 4380 a 6570 dias      60 a 100 bpm      50 a 90 bpm      12 a 20 rpm      < 90 mmHg

Fonte: SAVP (Pediatric Advanced Life Support - PALS), 2017a

**Tabela 6 – Tabelas de referências de exame de leucócitos (valores críticos)**

Faixa Etária	Frequência Cardíaca	
	Em vigília	Em sono
> 0 a 7 dias	100 a 205 bpm	90 a 160 bpm
> 7 dias a 28 dias	100 a 180 bpm	90 a 160 bpm
> 28 a 90 dias	98 a 140 bpm	80 a 120 bpm
> 90 a 365 dias	80 a 120 bpm	65 a 100 bpm
> 365 a 1825 dias	80 a 120 bpm	58 a 90 bpm
> 1825 a 4380 dias	75 a 118 bpm	58 a 90 bpm
> 4380 a 6570 dias	60 a 100 bpm	50 a 90 bpm

Fonte: Dados de referência do exame de Leucócitos do Laboratório do HOP

Para a limpeza dos registros, foi utilizado a ferramenta calc do pacote Libre Office, onde foi ordenado, todas as colunas respeitando a coluna de análise e quando identificado que o dado estava fora do padrão ou sendo um outlier, a linha do paciente inteira foi excluída.

Após a aplicação de técnicas de higienização, percebemos uma redução do Data Set de uma taxa de 10% da massa de dados.

Sendo, apresentado a tabela 7, a nova distribuição de totais de registros.

**Tabela 7 – Tabela de totais do datas higienizado:**

Faixa Etária	Pacientes com sinais de Sepsis	
	Não	Sim
> 0 a 7 dias	0	4
> 7 dias a 28 dias	15	14
> 28 a 90 dias	154	31
> 90 a 365 dias	359	69
> 365 a 1825 dias	2204	199
> 1825 a 4380 dias	1457	165
> 4380 a 6570 dias	3193	108
<b>Total Geral</b>	<b>7382</b>	<b>590</b>

Fonte: Autoria Própria

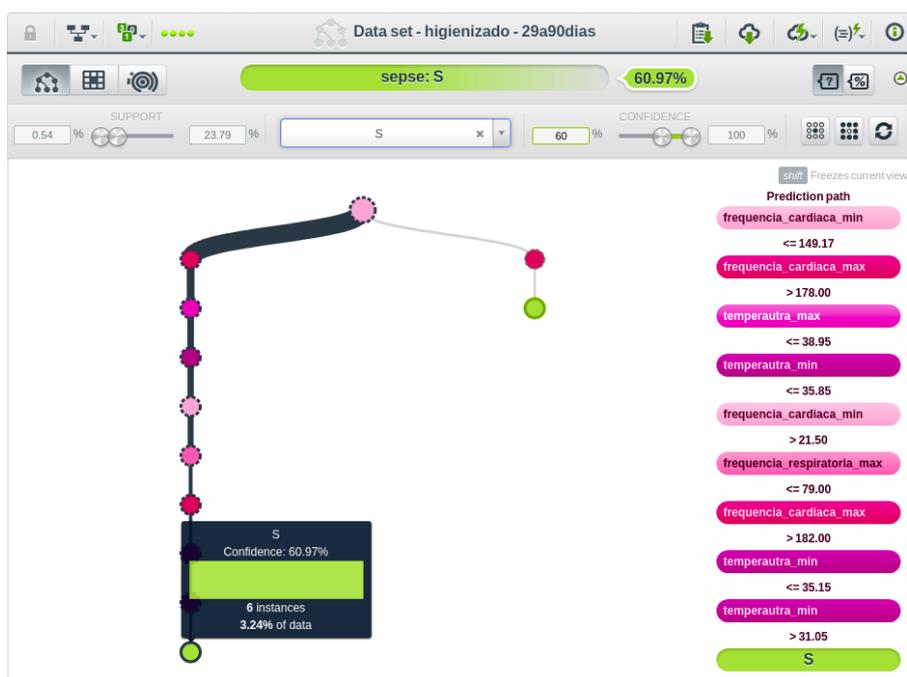
## 3.2 RESULTADOS

Com base na tabela apresentada após higienização, foi necessário desconsiderar a análise da(s) faixa(s) de crianças, de zero a sete dias e de sete a vinte oito dias de vida, pois seus dados não são representativos.

### 3.2.1 Resultado da análise de pacientes na faixa etária de 28 a 90 dias.

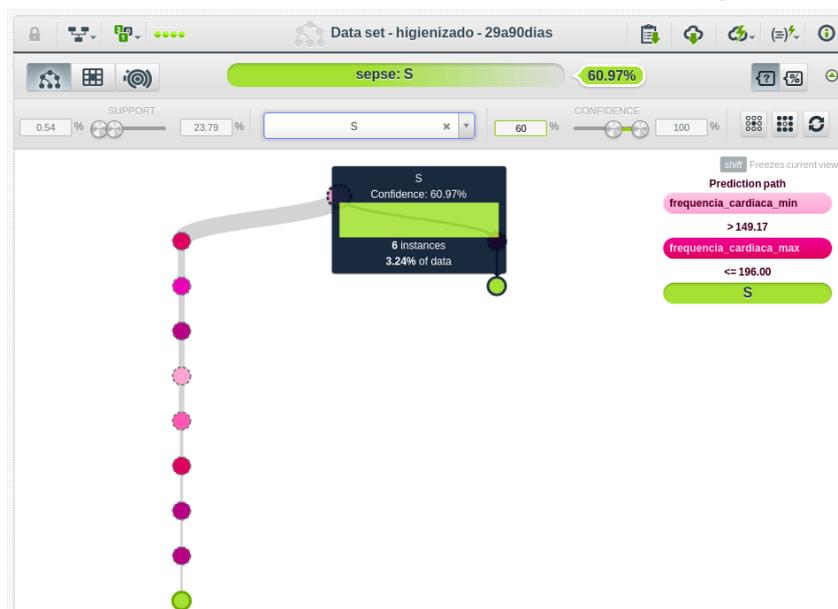
Ao definir a confiança de 60 a 100% é apresentado às árvores de decisão no gráfico 7:

**Gráfico 7 – Árvore de decisão com 60 a 100% de confiança – visão A**



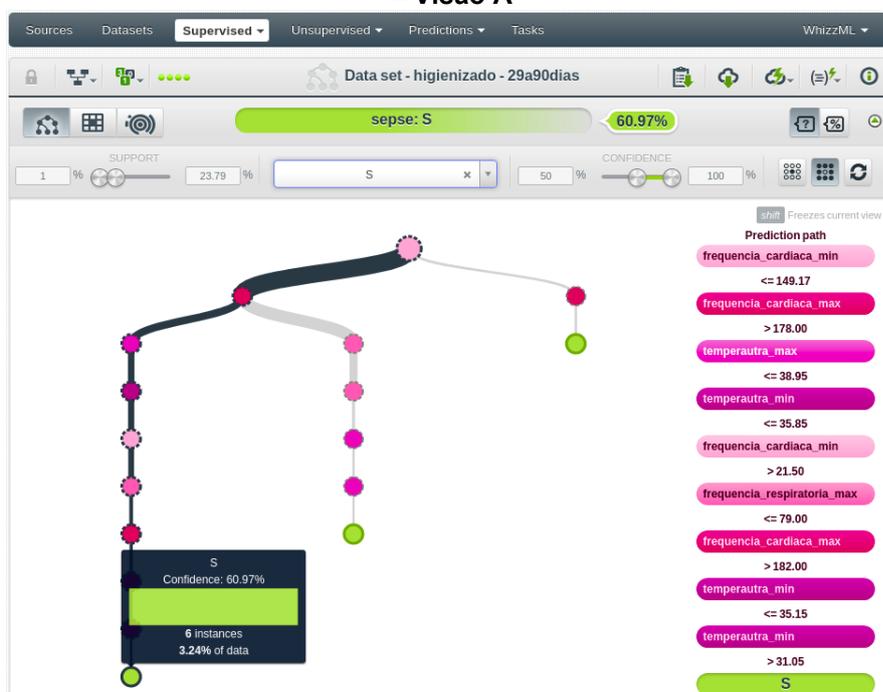
Fonte: Autoria Própria

Gráfico 8 – Árvore de decisão com 60 a 100% de confiança – visão B



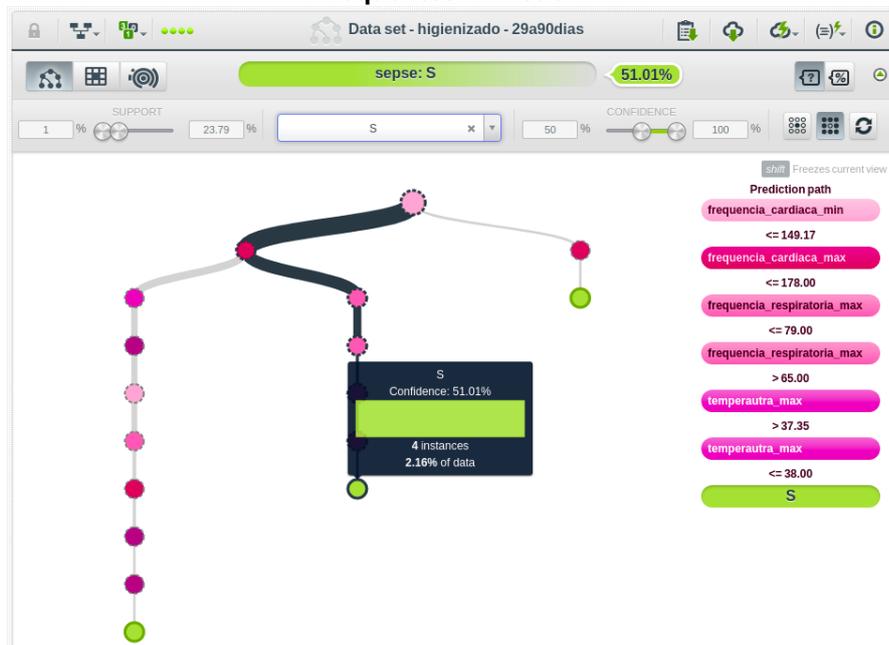
Fonte: Autoria Própria

Gráfico 9 – Árvore de decisão quando selecionado a opção 'padrões interessantes frequentes' – Visão A



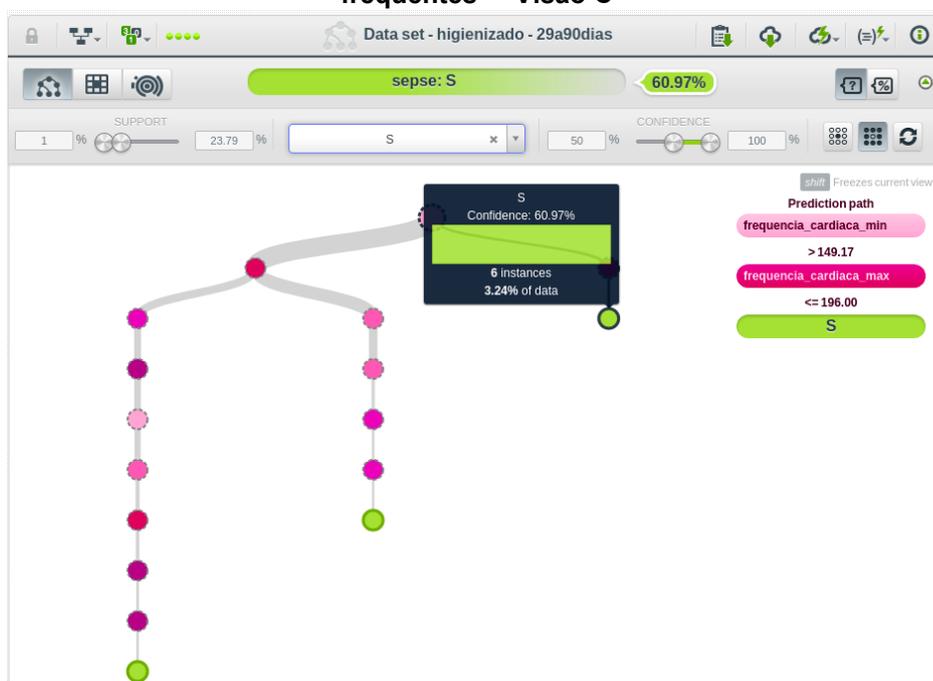
Fonte: Autoria Própria

**Gráfico 10 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão B**



Fonte: Autoria Própria

**Gráfico 11 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão C**

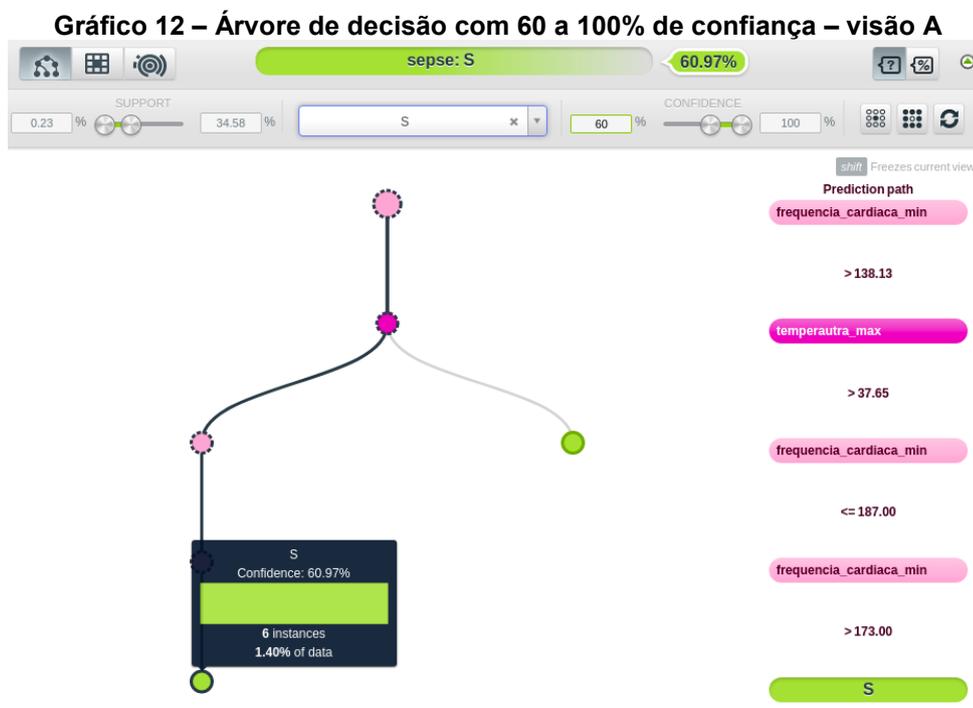


Fonte: Autoria Própria

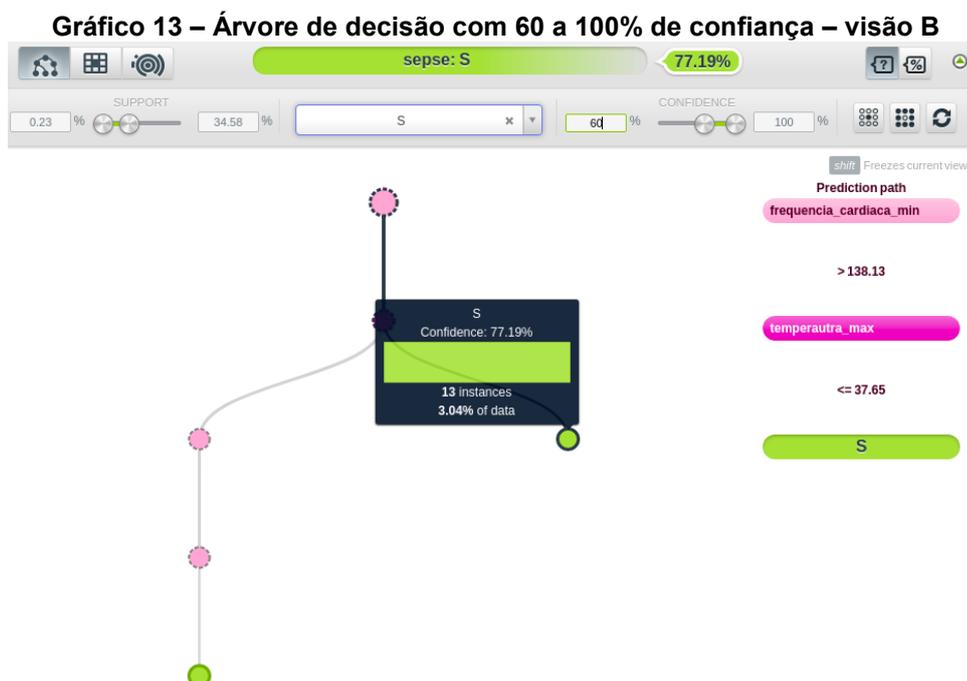
Com base nas árvores apresentadas da faixa de 29 a 90 dias, é possível perceber uma tendência em a frequência respiratória estar entre  $> 65$  e  $\leq 79$ , levando a uma leve mudança do padrão da tabela de parâmetros guia, e temperatura sendo apresentado entre  $> 37,35$  e  $\leq 38$  reforçando a tabela SIRS.

### 3.2.2 Resultado da análise de pacientes na faixa etária de 90 a 365 dias de vida.

Ao definir a confiança de 60 a 100% é apresentado às árvores de decisão no gráfico 12:

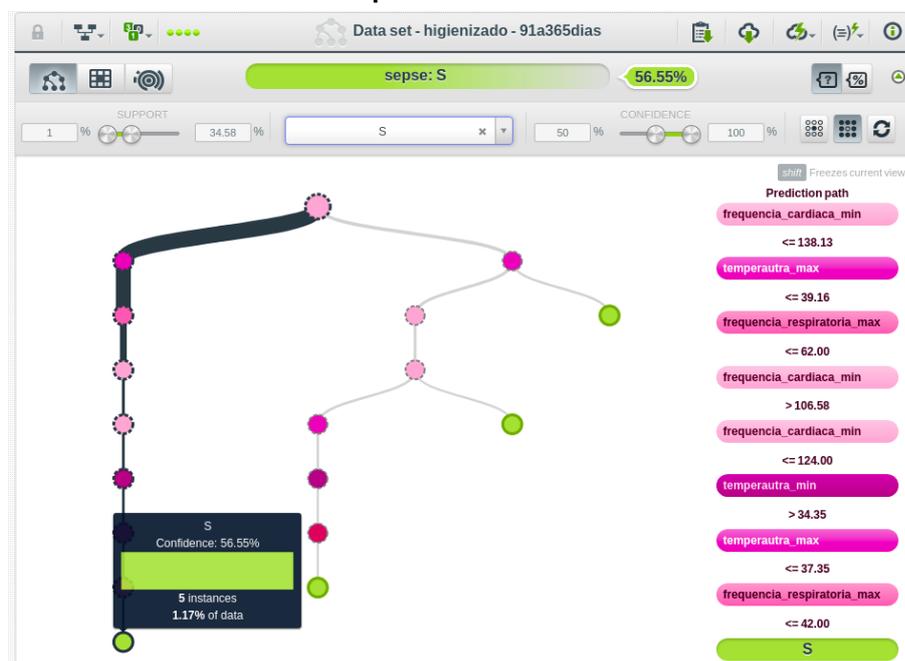


Fonte: Autoria Própria



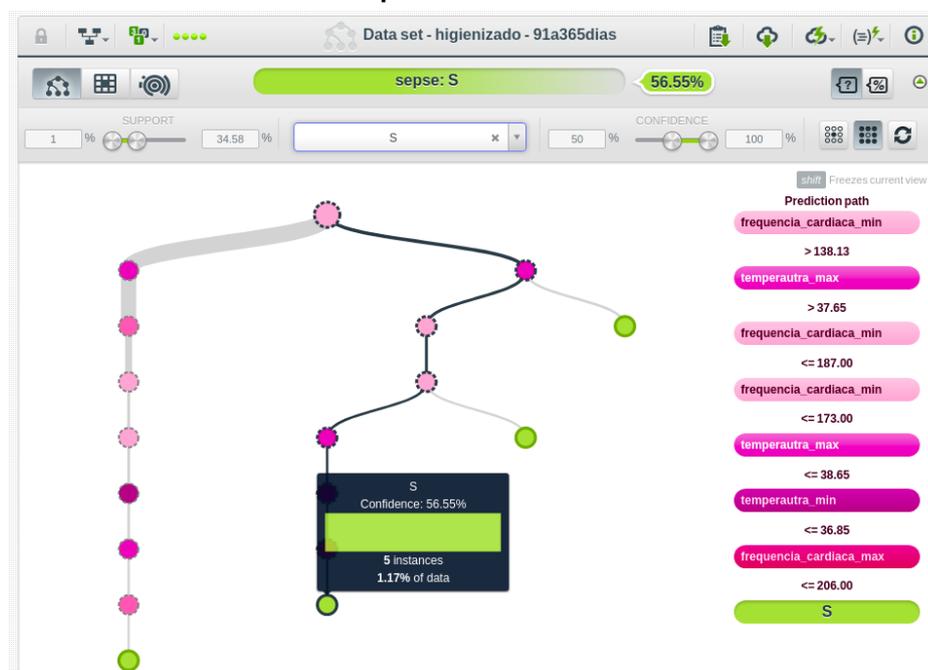
Fonte: Autoria Própria

**Gráfico 14 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão A**



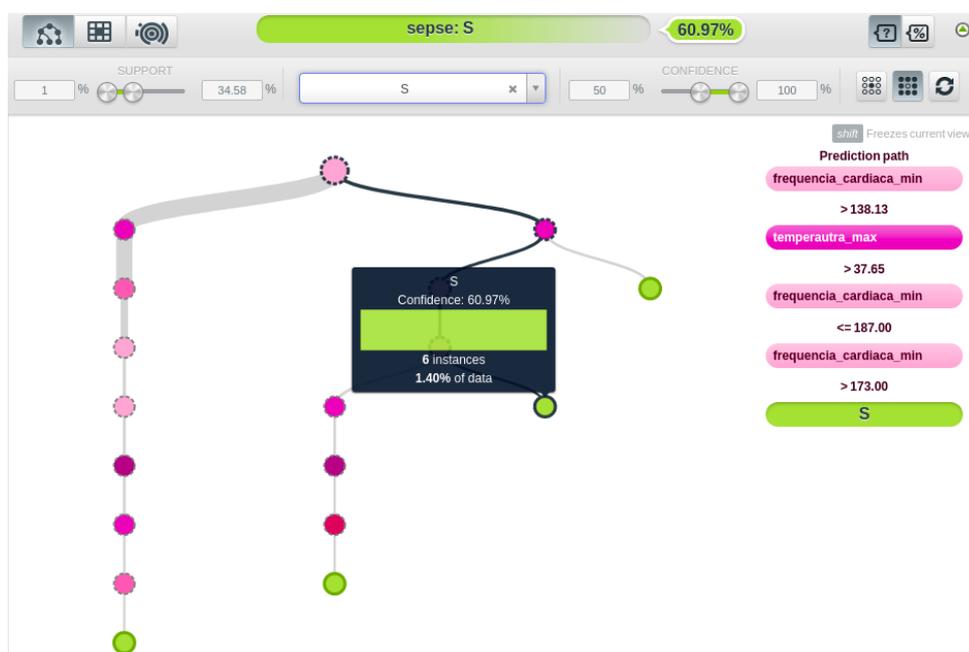
Fonte: Autoria Própria

**Gráfico 15 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão B**



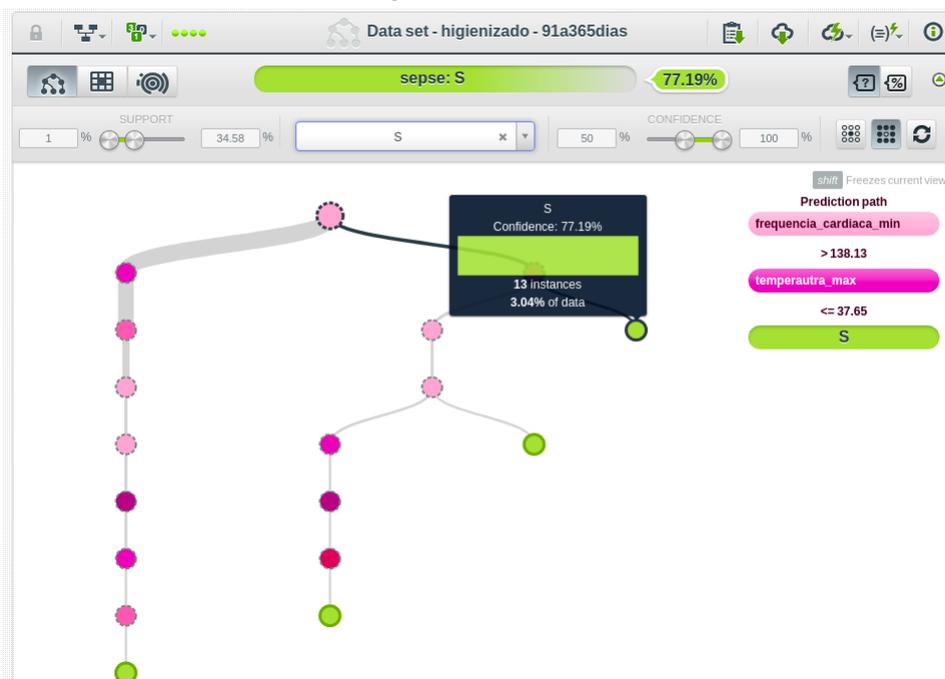
Fonte: Autoria Própria

Gráfico 16 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão C



Fonte: Autoria Própria

Gráfico 17 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’ – Visão D



Fonte: Autoria Própria

Com base nas árvores apresentadas da faixa de 90 a 365 dias, embora as árvores tenham demonstrado mais visões, não foi possível perceber nenhuma informação que seja relevante.

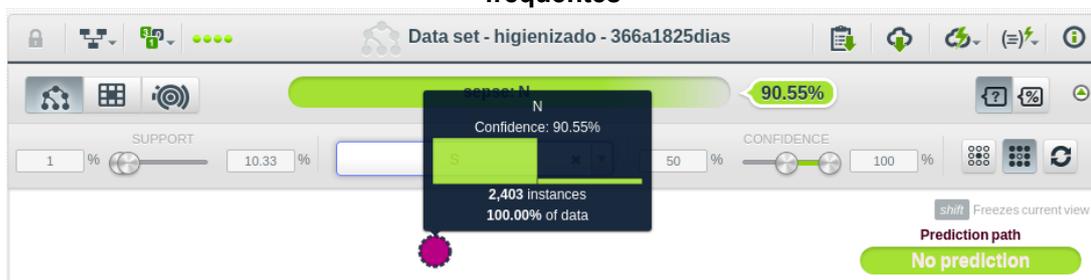
### 3.2.3 Resultado da análise de pacientes na faixa etária de 366 a 1825 dias.

Ao definir a confiança de 60 a 100% é apresentado às árvores de decisão no gráfico 18:



Fonte: Autoria Própria

**Gráfico 19 – Árvore de decisão quando selecionado a opção 'padrões interessantes frequentes'**



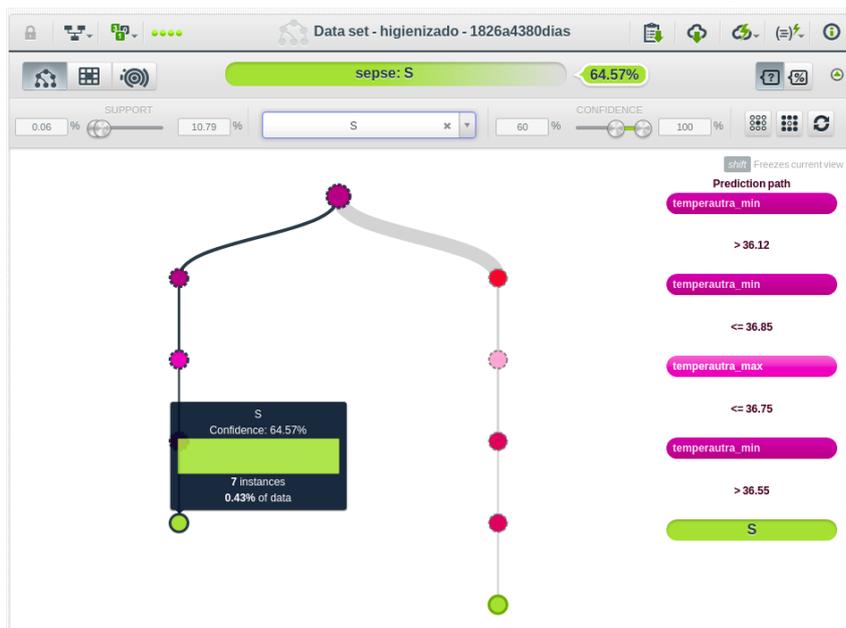
Fonte: Autoria Própria

Com base nas árvores apresentadas da faixa de 366 a 1825 dias, embora a amostra seja maior, a árvore não conseguiram apresentar um padrão e como é demonstrado no gráfico 20, que não é possível obter predição com base nos padrões otidos.

### 3.2.4 Resultado da análise de pacientes na faixa etária de 1825 a 4380 dias.

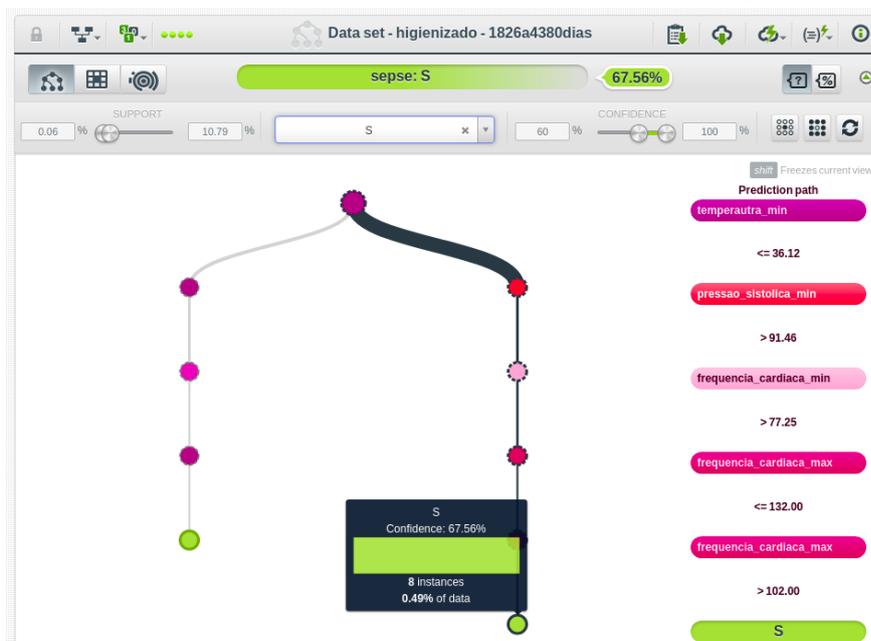
Ao definir a confiança de 60 a 100% é apresentado às árvores de decisão no gráfico 20:

**Gráfico 20 – Árvore de decisão com 60 a 100% de confiança – visão A**



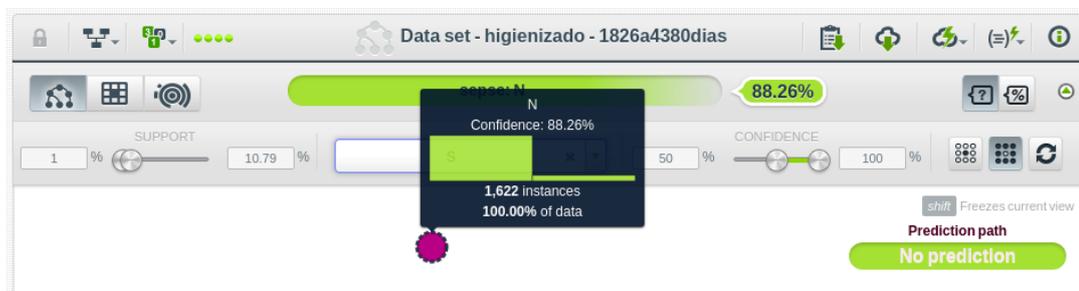
Fonte: Autoria Própria

**Gráfico 21 – Árvore de decisão com 60 a 100% de confiança – visão B**



Fonte: Autoria Própria

**Gráfico 22 – Árvore de decisão quando selecionado a opção ‘padrões interessantes frequentes’**



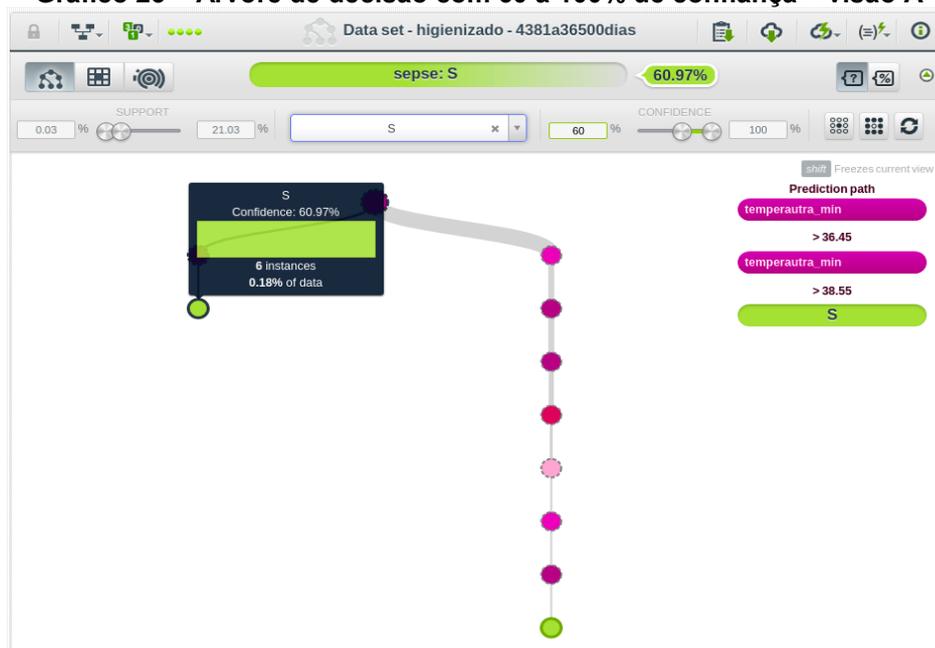
Fonte: Autoria Própria

Com base nas árvores apresentadas da faixa de 1826 a 4380 dias, foi possível observar que a árvore não gerou um padrão e como demonstrado no no gráfico 23, não é possível obter predição.

### 3.2.5 Resultado da análise de pacientes na faixa etária de 4380 a 6570 dias.

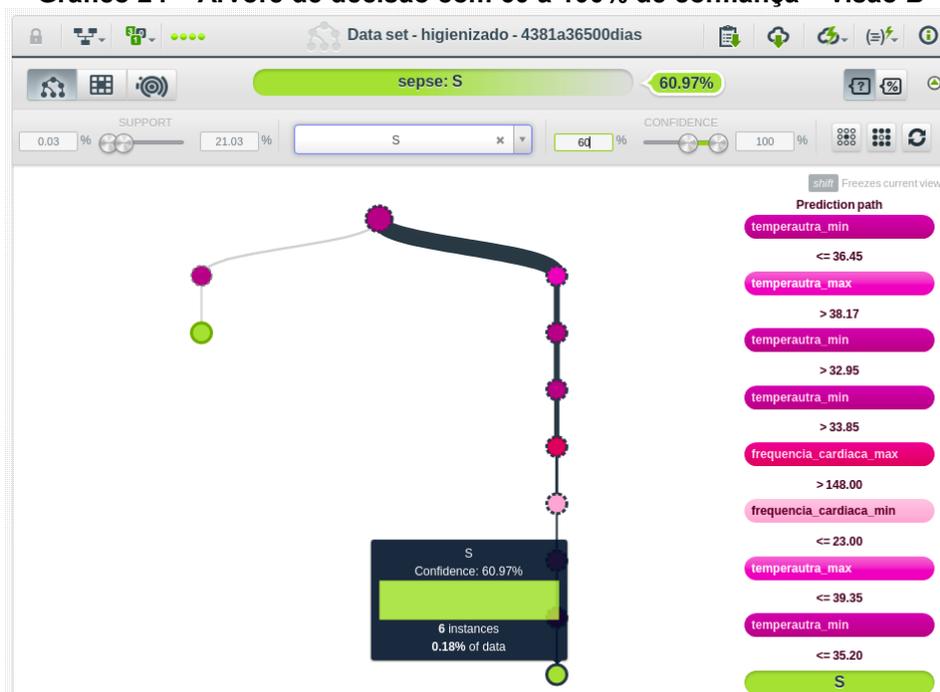
Ao definir a confiança de 60 a 100% é apresentado às arvores de decisão no gráfico 23:

**Gráfico 23 – Árvore de decisão com 60 a 100% de confiança – visão A**



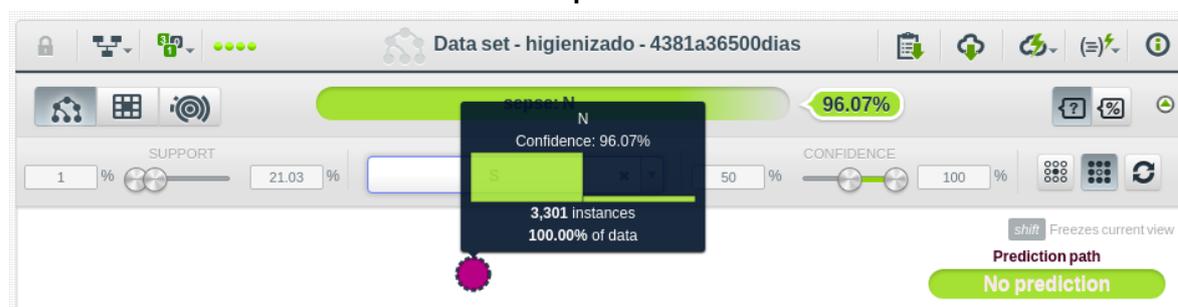
Fonte: Autoria Própria

Gráfico 24 – Árvore de decisão com 60 a 100% de confiança – visão B



Fonte: Autoria Própria

Gráfico 25 – Árvore de decisão quando selecionado a opção 'padrões interessantes frequentes'



Fonte: Autoria Própria

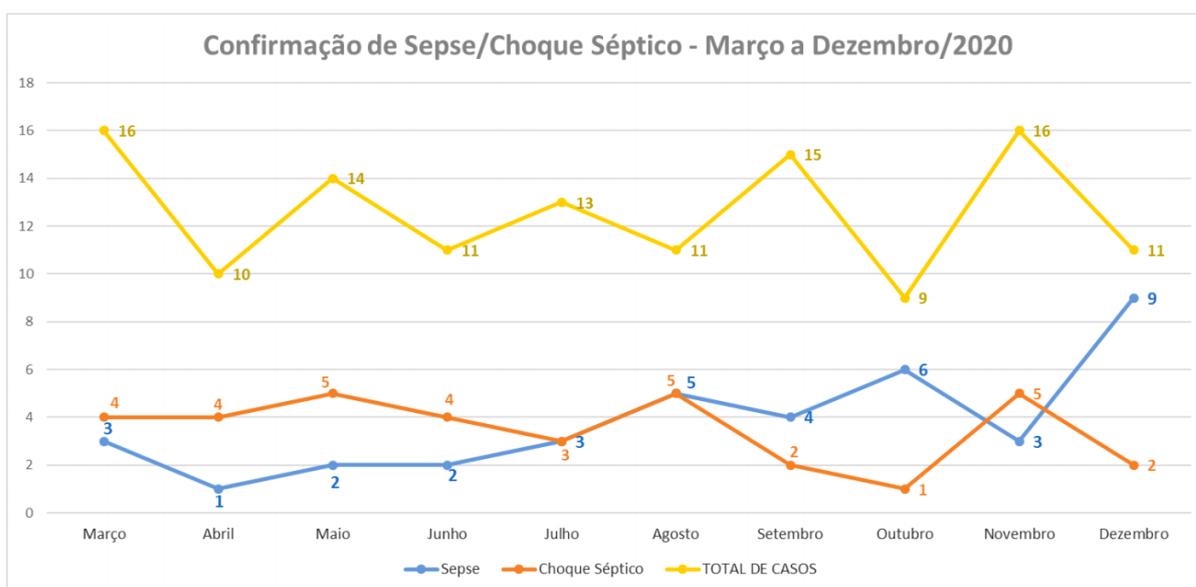
Com base nas árvores apresentadas da faixa de 4381 a 36500 dias, foi possível perceber que a árvore reforça o dado estabelecido na tabela SIRS, sendo temperature estando a cima de 38 graus, na arvore é demonstrado um intervalo sendo maior que 38,17 e menor, igual a 39,35. Mesmo assim conforme demonstrado no gráfico 23, não é possível obter predição com base nos padrões otidos.

### 3.3 DADOS DO HOSPITAL OBJETO DA PESQUISA (HOP)

O hospital HOP é um dos maiores de Curitiba e referência nacional na assistência a saúde. Uma instituição filantrópica, desde sua fundação, olha com

carinho para à causa da saúde infanto-juvenil. No HOP, são realizados por ano mais de 288 mil atendimentos ambulatoriais e oferece mais de 30 especialidades médicas, com aproximadamente 400 leitos de internação, sendo aproximadamente 80 em UTIs, realizando mais de 20 mil cirurgias por ano. Cerca de 70% da sua capacidade de atendimento é destinada para o Sistema Único de Saúde (SUS). Isso faz com que receba pacientes de diferentes estados do Brasil.

**Gráfico 26 - Gráfico com o número de infecções por sepse do HOP.**



**Fonte: Departamento do Núcleo da Qualidade do (HOP)**

#### **4 CONCLUSÃO (OU CONSIDERAÇÕES FINAIS)**

Aplicar algoritmo de Machine learning sobre dados de sinais vitais e resultados de exame de laboratório de pacientes que tiveram alertas de sepse positivos e negativos foi muito interessante e de grande ganho no aprendizado, pois conseguimos observar padrões. O trabalho teve o objetivo de reavaliar os parâmetros da tabela SIRS com base em dados reais, e com esta visão, tivemos algumas reafirmações na teoria aplicada pela tabela SIRS, mas também várias outras informações sem relevância. Porém, embora na totalização dos registros, após a higienização serem números relevantes, quando fracionado em faixas etárias, se torna uma amostra irrelevante, e por conta disto, chegou-se à conclusão que esta pesquisa não alcançou com plenitude o objetivo que se buscava.

Porém, este trabalho serviu como base e deverá ser utilizado para o início de uma nova linha de pesquisa, talvez uma linha de pesquisa de mestrado, pois foi possível perceber que podemos cruzar dezenas de outras informações, que incorporada à análise, novos parâmetros poderão ser acrescentados a tabela SIRS e naturalmente, contribuir para o diagnóstico precoce da Sepse e por sua vez, salvar vidas.

## REFERÊNCIAS

SCOTT L. Weiss; Diretrizes internacionais da Campanha de Sobrevivência à Sepse para o tratamento de choque séptico e disfunção orgânica associada à sepse em crianças, *Pediatric Critical Care Medicine*, 2020, disponível em: <https://doi.org/10.1097/PCC.0000000000002198>), Acesso em: 05/06/2021

ILAS (Instituto Latino Americano da Sepse). **Sepsis**. Disponível em: <<https://www.ilas.org.br/materiais-pediatria.php>>. Acesso em 04/07/2021.

GRUS, Joel **Data Science do Zero**: Primeiras regras com o Python, 2016.

MITCHELL, Tom M. *Machine Learning*. 1ª Edição, 1997.

Rosa, João L.G. **Fundamento da Inteligência Artificial**, 2011.

ILAS (Instituto Latino Americano de Sepse). **Protocolo Clínico**: Implementação de protocolo gerenciado de sepse, 2018.

Provost, F., Fawcett.T. **Data Science para Negócios**: O que Você Precisa Saber Sobre Mineração de Dados e Pensamento Analítico de Dados,2016.

Cellier P., Driessens K. **Machine Learning and Knowledge Discovery in Databases**. 1996.