

LETÍCIA MELCOP SANT'ANNA

Avaliação da Aderência e dos Desvios de Fluxo em um Departamento de Emergência: Uma Aplicação de Mineração de Processos

**PROJETO DE GRADUAÇÃO EM ENGENHARIA DE PRODUÇÃO
APRESENTADO AO DEPARTAMENTO DE ENGENHARIA INDUSTRIAL
DA PUC-RIO, COMO PARTE DOS REQUISITOS PARA OBTENÇÃO
DO TÍTULO DE ENGENHEIRO DE PRODUÇÃO**

**Orientador: Leonardo dos Santos Lourenço Bastos
Coorientador: Soraida Aguilar**

**Departamento de Engenharia Industrial
Rio de Janeiro, 24 de novembro de 2025.**

Agradecimentos

Primeiramente gostaria de agradecer aos meus orientadores, Leonardo e Soraida, que estiveram comigo durante todo o processo. Não poderia pedir por melhores professores me acompanhando nessa jornada. Quero agradecer também à minha família: à minha mãe, que me abraçou sempre que eu precisei, ao meu pai, o qual eu só quero trazer orgulho, e ao meu irmão que sempre me deixava gritar no quarto dele quando eu precisava extravasar e sempre cuidava de mim quando eu mesma não conseguia.

Para os meus amigos que me deram forças: vocês vivem em meu coração. Luiza e Matheus que passaram longas horas trabalhando junto comigo, Giovana e Paula que acreditaram em mim ainda que eu não acreditasse, Rafaela e Marina que surtaram igual a mim com o final da faculdade, e os amigos do vôlei e da RioBotz que me trouxeram paz em momentos que eu nem sabia que precisava.

Obrigada a todas as pessoas que tiveram uma mão no meu caminho até então, seja por uma hora ou por uma década, eu não estaria aqui de outra maneira.

Resumo

O gerenciamento de processos é fundamental para melhorar desempenho e qualidade de processos organizacionais, especialmente na área da saúde, que tem processos caracterizados por alta complexidade e variabilidade. Essas questões são agravadas ainda mais nos Departamentos de Emergência (DE), cuja operação envolve decisões críticas em tempo reduzido e fluxos de tratamento diversos. Neste estudo, aplicou-se a Mineração de Processos, seguindo um ciclo adaptado de ciência de dados, para comparar o fluxo teórico definido por um hospital privado brasileiro com o fluxo real registrado em seu sistema, avaliando a aderência por meio da métrica de *fitness* e investigando fatores que explicam desvios, como idade, diagnóstico e gravidade. Os resultados mostram que, embora 1/3 dos pacientes sigam as etapas esperadas, os desvios de caminho ocorrem em casos mais graves, com fluxos mais longos e complexos. Concomitantemente, inconsistências de registro também contribuem para parte das divergências observadas em relação ao fluxo teórico. Conclui-se que a Mineração de Processos oferece uma visão clara dos fluxos reais, evidenciando como características clínicas e a qualidade dos dados influenciam a aderência ao processo e apontando oportunidades de melhoria operacional no DE.

Palavras-chave

Mineração de Processos, Gerenciamento de Processos, Departamento de Emergência

Título em inglês

Assessment of Process Compliance and Flow Deviations in an Emergency Department:
A Process Mining Application

Abstract

Process management is essential for improving the performance and quality of organizational processes, especially in healthcare, where processes are characterized by high complexity and variability. These challenges become even more problematic in Emergency Departments, where operations involve time-critical decisions and diverse treatment paths. In this study, Process Mining was applied, following an adapted data-science lifecycle, to compare the theoretical process model defined by a Brazilian private hospital with the actual patient flow recorded in its system, assessing compliance through the fitness metric and examining factors that explain deviations, such as age, diagnosis, and severity of the case. The results show that although 1/3 of patients follow the expected sequence of steps, path deviations occur among more severe cases, which exhibit longer and more complex trajectories. Additionally, system inconsistencies also contribute to some of the divergences observed in relation to the theoretical model. The study concludes that Process Mining provides a clear view of real patient flows, revealing how clinical characteristics and data quality influence process adherence and highlighting opportunities for operational improvement in emergency departments.

Key-words

Process Mining, Process Management, Emergency Department

Sumário

1. Introdução	1
2. Referencial Teórico.....	4
2.1. Modelagem de Processos.....	4
2.1.1. <i>Business Process Model and Notation</i>	4
2.1.2. Rede de Petri (<i>Petri Net</i>)	5
2.1.3. Modelo <i>Directly Follows</i>	6
2.2. <i>Process Mining</i> (Mineração de Processos)	7
2.2.1. Algoritmos de <i>Process Discovery</i>	8
2.2.2. Métodos de Avaliação de Modelos.....	10
2.3. Ciclo de Vida de Ciência de Dados.....	12
2.4. Departamento de Emergência	14
3. Materiais e Métodos	16
3.1. Entendimento do Problema.....	16
3.2. Coletar e Gerenciar Dados	18
3.3. Construir Modelo e Avaliar Modelo	20
4. Resultados e Discussão.....	23
4.1. Análise Descritiva	23
4.2. Análise de Conformidade.....	28
4.3. <i>Process Discovery</i>	31
4.4. Considerações Adicionais	36
5. Conclusão	38
6. Bibliografia.....	40

Lista de Figuras

Figura 1 - Exemplos de notações do BPMN	5
Figura 2 - Exemplo de uma <i>Petri Net</i>	6
Figura 3 - Exemplo de Modelo Directly Follows.....	7
Figura 4 - Padrões típicos do Algoritmo Alpha e sua representação em uma Petri Net...	9
Figura 5 - Exemplos de <i>Fitness</i>	12
Figura 6 - Ciclo de Vida de Ciência de Dados	13
Figura 7 - Algoritmo de fluxo do paciente no Serviço Hospitalar de Urgência e Emergência.	14
Figura 8 - Modelo BPMN do Fluxo de Pacientes no Departamento de Emergência do Hospital em Questão.....	17
Figura 9 - Modelo Directly Follows do Processo do Departamento de Emergência	21
Figura 10 - <i>Boxplot</i> de Distribuição de Idades por Categoria de Gravidade.....	24
Figura 11 – Volume Acumulado de Pacientes no Departamento de Emergência por Mês e Ano.	25
Figura 12 – Volume Acumulado de Pacientes no Departamento de Emergência por Dia da Semana	26
Figura 13 – Volume Acumulado de Pacientes no Departamento de Emergência por Hora de Entrada	27
Figura 14 – Volume Acumulado de Pacientes no Departamento de Emergência por Turnos de Entrada e Dia da Semana	27
Figura 15 – Porcentagem de Níveis de Gravidades no Departamento de Emergência por Hora de Entrada	28
Figura 16 - <i>Fitness</i> do Trace de cada Paciente	29
Figura 17 - Modelo Minerado do Grupo A	31
Figura 18 - Modelo Minerado do Grupo B	32
Figura 19 - Modelo Minerado do Grupo C	33
Figura 20 - Modelo Minerado dos Pacientes Categoria Azul.....	34
Figura 21 - Modelo Minerado dos Pacientes Categoria Verde.....	34
Figura 22 - Modelo Minerado de Pacientes Categoria Amarelo	35
Figura 23 - Modelo Minerado dos Pacientes Categoria Laranja e Vermelho.....	36

Lista de Tabelas

Tabela 1 - Sistema Manchester de Triagem.....	15
Tabela 2 - Horários de Plantão da Equipe Médica	20
Tabela 3 - Quantidade de Médicos por Turno e Dia da Semana no Departamento de Emergência	20
Tabela 4 - Tabela Descritiva dos Pacientes.....	23
Tabela 5 - Tabela Descritiva por <i>Fitness</i>	29
Tabela 6 - Porcentagem de Pacientes Atendidos a Tempo por Gravidez.....	37

1. Introdução

Processos de diversos tipos compõem o dia a dia de todas as organizações – as sequências de eventos, atividades e decisões que trazem valor ao cliente final ou à organização. Tradicionalmente, o gerenciamento de processos (BPM – *Business Process Management*) consiste em um processo cíclico de modelagem, análise, redesenho e implementação, buscando trazer benefícios de performance, experiência ou custo para as organizações e clientes (Dumas et al., 2018).

Algumas metodologias mais usadas na gestão de processos, como Análise de Filas, Análise de Causa-Raíz, Quadrilátero do Diabo (ou "Devil's Quadrangle"), auxiliam nas análises e redesenho, baseando-se em modelagens do processo “*AS-IS*” (como é, em tradução livre), construídos a partir de contato direto com os executores e administradores do processo (Dumas et al., 2018). Contudo, os processos desenhados podem não ser condizentes com a realidade na qual eles ocorrem, impedindo a aplicação adequada das ferramentas de análise de processos (Van der Aalst, 2016). Dessa forma, a modelagem de processos de maneira precisa é essencial, não apenas para entender o funcionamento atual do processo, mas também para possibilitar a aplicação de metodologias que revelam desvios, gargalos e oportunidades de melhoria.

No contexto dos processos organizacionais, o sistema de saúde se destaca por ser uma área caracterizada por processos complexos, com alta variabilidade de fluxos e fatores contextuais e patológicos (Dallagassa et al., 2022). Dentro desse sistema, o departamento de emergência (DE) é um dos ambientes com muitos desafios, devido a criticidade dos pacientes que entram e a necessidade do pronto-atendimento.

Os principais desafios operacionais dentro de um DE estão relacionados a: superlotação, longos tempos de espera, estadia prolongada, excessivo tempo de processamento dos pacientes e um número elevado de pacientes que deixam o serviço sem atendimento (Ortíz-Barrios; Alfaro-Saíz, 2020). Problemas estes, que na literatura são tratados muitas vezes com ferramentas relacionadas à modelagem de processos e técnicas BPM como simulação e abordagens *lean* (Ortíz-Barrios; Alfaro-Saíz, 2020).

Com o crescimento da transformação digital nas organizações, se tornou possível registrar os eventos e atividades dos processos de forma sistemática, trazendo assim uma maior visibilidade da ocorrência real dos processos (Van der Aalst, 2016). Dessa forma, a disciplina de *Process Mining* (Mineração de Processos) surge como o conjunto de

técnicas capaz de avaliar e extrair informações dos sistemas, trazendo uma visão factual para o BPM (Van der Aalst; Dustdar, 2012).

A literatura em Mineração de Processos tem uma parte significativa de atuação em estudos na área de saúde (Garcia et al., 2019). No Brasil, a literatura também tem contribuições significativas, como os estudos de Dallagassa et al. (2022) sobre oportunidades e desafios da aplicação, de Neira et al. (2019) sobre os usos de Mineração de Processos em um hospital terciário, e de Pereira, Santos e Maceno (2020) sobre o fluxo de tratamento de sépsis.

Embora a Mineração de Processos venha sendo aplicado em DE's para a descoberta de processos, como em Pegoraro et al. (2023), e para a melhoria de desempenho, como em Pereira, Santos e Maceno (2020), existe ainda uma lacuna na compreensão da aderência real dos processos idealizados em relação aos processos que ocorrem durante o dia a dia. Com isso, as seguintes perguntas de pesquisa surgem em relação a essa problemática:

PP1 – Quais são as diferenças principais entre um fluxo real de pacientes em um departamento de emergência de um hospital quando comparado ao modelo de fluxo de referência?

PP2 – Quais características de pacientes e organização estão relacionadas a desvios no processo dentro de um departamento de emergência?

PP3 – Quais fluxos devem ser considerados de acordo com os determinantes do processo de maneira robusta e eficiente?

Destarte, o objetivo desse trabalho é analisar o fluxo real registrado sistematicamente com o fluxo desenhado pelo hospital, procurando explicitar quais fatores causam desvios, como diagnóstico, idade, gravidade da situação e outros. Para isso, será utilizada a base histórica de um departamento de emergência de um hospital privado brasileiro.

Por fim, a organização deste estudo segue a seguinte estrutura. A seção inicial, Capítulo 1, Introdução, detalha o contexto e a importância do assunto em evidência, além de incluir as questões investigativas e os objetivos do trabalho. O Capítulo 2 trata do Referencial Teórico, que explora os fundamentos do tema. Em seguida, o Capítulo 3 apresenta a metodologia utilizada para a obtenção dos resultados, baseada no ciclo de

vida de um projeto de ciência de dados. No Capítulo 4, são apresentados e discutidos os resultados obtidos em relação às perguntas de pesquisa. Finalmente, o Capítulo 5 conclui o trabalho, resumindo as principais informações e levantando ideias para trabalhos futuros.

2. Referencial Teórico

Neste capítulo do trabalho, serão explicitados os conceitos teóricos para a compreensão e desenvoltura do projeto. Ele será dividido em quatro seções: Modelagem de Processos, apresentando ferramentas e ideias gerais; *Process Mining* (Mineração de Processos), aprofundando-se nos termos e fundamentos; Ciclo de Vida de Ciência de Dados, detalhando o framework para análises envolvendo dados; e por fim, Departamentos de Emergências, com uma explicação sobre o funcionamento desse setor em hospitais.

2.1. Modelagem de Processos

A gestão de processos busca estruturar, monitorar e melhorar fluxos de trabalho organizacionais, garantindo que atividades sejam executadas de maneira eficiente, padronizada e orientada a objetivos claros (Dumas et al., 2018). Dessa forma, a representação de um processo é essencial para o entendimento completo dos fatores e atividades envolvidas durante sua execução. A modelagem é a maneira de conseguir essa representação visual, porém, existem diversas notações diferentes para desenhar os processos. Nesse trabalho, serão explicitadas as modelagens BPMN, Petri Net e *Directly Follows*.

2.1.1. Business Process Model and Notation

O *Business Process Model and Notation*, ou BPMN, tem como objetivo principal ser uma notação padronizada para modelagem de processos que possa ser compreendida por todos. Sua padronização facilita a comunicação entre organizações e profissionais, reduz a necessidade de compreender múltiplas metodologias e melhora o entendimento de processos ao longo de seu ciclo do BPM. O BPMN também evolui em relação às notações tradicionais ao suportar conceitos avançados, como processos com exceções, transações e decisões, tornando-se uma ferramenta mais completa e flexível para modelar e gerir processos de negócio (Object Management Group, 2010).

A linguagem BPMN possui um vocabulário composto por atividades, representadas por caixas arredondadas, eventos, representados por círculos; gateways,

representados por losango; e fluxos de sequência, representados por setas. Todo este vocabulário está regido por regras sintáticas, como a necessidade de que todos os elementos estejam conectados do nó de início ao nó de fim e que toda atividade tenha apenas um fluxo de entrada e um de saída (Dumas et al., 2018).

Na Figura 1, encontra-se exemplos de algumas notações mais comuns no BPMN como o evento de início de processo, evento de fim, *gateway* exclusivo – o qual representa que apenas um caminho é seguido, *gateway* paralelo – o qual representa que todos os caminhos são seguidos simultaneamente, a caixa de atividade, e o conector entre as notações.



Figura 1 - Exemplos de notações do BPMN. Adaptado de Object Management Group (2010).

2.1.2. Rede de Petri (*Petri Net*)

Uma Rede de Petri, ou *Petri Net*, é uma modelagem de processos útil para sistemas de processamento de informação, composta por três elementos: *Lugares*, *Transições* e *Tokens* (Figura 2). *Transições* são representadas por quadrados e conectadas a *Lugares*, que são simbolizados por círculos. *Tokens*, por sua vez, são representados por pequenos círculos, existem dentro dos *Lugares* e habilitam *Transições*: uma *Transição* só pode ocorrer, quando todos os seus *Lugares* de entrada contêm um *Token*. A dinâmica de uma *Petri Net*, ocorre quando uma *Transição* dispara, consumindo *Tokens* de todos os *Lugares* conectados entrando nela e produzindo *Tokens* em *Lugares* conectados saindo da *Transição*, permitindo representar um fluxo de estados e atividades (Van der Aalst, 2016).

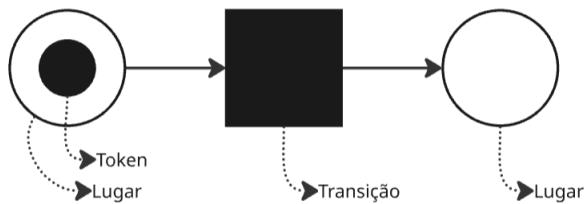


Figura 2 - Exemplo de uma *Petri Net*. Fonte: Elaborado pela Autora.

A modelagem de processos de negócio em *Petri Net* se configura em uma *Workflow Net*, uma subclasse de *Petri Net* com a condição de *soundness*, que engloba as seguintes restrições (Van der Aalst, 2016):

- *Safeness* (segurança): Um lugar não pode ter nunca mais de um token;
- *Proper Completion* (completude): Deve existir uma circunstância na qual não há tokens em nenhum lugar que não seja o lugar final;
- *Option to Complete* (possibilidade de concluir): Todos os tokens são capazes de chegar no lugar final por meio de transições;
- *Absence of Dead Parts* (ausência de partes “mortas”): Não existem transições que não possam ser ativadas.

Além disso, *Workflow Nets* se alinham efetivamente à Mineração de Processos, já que suas sequências de disparo correspondem àquelas registradas em sistemas informacionais (Van der Aalst, 2016).

2.1.3. Modelo *Directly Follows*

O Modelo *Directly Follows* tem como foco principal organizar a sequência em que as atividades ocorrem. Essa modelagem consiste, essencialmente, em um grafo direcionado, cujos nós representam atividades e as setas representam a transição de uma atividade para outra, trazendo uma visão simples da sequência em que atividades ocorrem (Leemans; Poppe; Wynn, 2019).

Assim como o *Workflow Net*, o modelo *Directly Follow* deve possuir a característica de *soundness*: nesse contexto isso significa principalmente que qualquer nó está em um caminho que vai do nó de início (o qual não possui nenhuma transição entrando nele) ao nó de fim (o qual não possui nenhuma transição saindo dele) (Leemans; Poppe; Wynn, 2019). Na Figura 3, é possível ver um exemplo simples de um modelo *Directly Follows*, no qual apenas duas sequências são possíveis: início, A, B, C, E, fim ou início, A, B, D, E, fim.

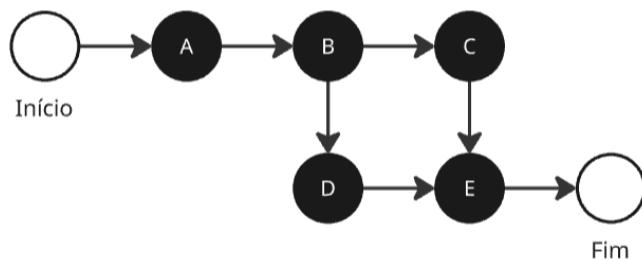


Figura 3 - Exemplo de Modelo Directly Follows. Fonte: Elaborado pela Autora.

2.2. Process Mining (Mineração de Processos)

Mineração de Processos é o elo que faz a conexão entre conhecimentos de BPM e os dados gerados pelos processos. A premissa fundamental para a execução de Mineração de Processos é o fato de que o processo a ser estudado tenha seus eventos registrados em um sistema de maneira sequencial e que seja possível extrair essas informações, gerando *eventlogs* (Van der Aalst; Dustdar, 2012).

Um *eventlog* é uma coleção de *traces* (rastros), que por sua vez é uma coleção de eventos (Diba et al., 2020). Um evento está relacionado a alguma atividade do processo e tem no mínimo dois atributos essenciais atrelados: o *case id* (caso): uma chave única de identificação da instância, como por exemplo, CPF de um paciente; e um *timestamp*: data e hora referente ao início da atividade (Van der Aalst, 2016). Outros atributos também podem estar atrelados ao evento como duração, recurso utilizado, custo etc., trazendo mais informações para as análises (Van der Aalst; Dustdar, 2012).

Dessa forma, um *trace* é a união de todos os eventos referentes a um caso, organizados de forma cronológica, mapeando assim o caminho percorrido pela entidade

do processo. Consequentemente, o *eventlog* é composto por todos os *traces* registrados em um determinado período (Diba et al., 2020).

Com o *eventlog* e atributos dos casos, a Mineração de Processos pode ser utilizada para diversos benefícios organizacionais, como identificação de gargalos, ineficiências, utilização de recursos, regras de decisão e outros (Van der Aalst, 2016). A disciplina está dividida atualmente em três principais modalidades (Van der Aalst; Dustdar, 2012):

- *Process Discovery* (Descoberta de Processo): cria um modelo de processo com base apenas em um *eventlog*. Isso pode ser realizado com diversas técnicas, algoritmos e softwares diferentes;
- *Conformance Checking* (Verificação de Conformidade): compara o alinhamento entre *eventlog* e modelo de processo. Diferentes métricas podem ser utilizadas, como *fitness* (aderência) e precisão (Carmona; Van Dongen; Weidlich, 2022), para avaliar o quanto o modelo difere da base de dados ou vice-versa;
- *Performance Enhancement* (Aprimoramento de Performance): utiliza o *eventlog* em conjunto com um modelo de processo para melhorar a performance de um processo, procurando obter ganhos em nível de serviço, tempo de processamento ou eliminar gargalos.

2.2.1. Algoritmos de *Process Discovery*

Através de algoritmos de *Process Discovery*, é possível construir modelos de processos com base em *eventlogs*. Nesse trabalho será ressaltado o funcionamento de duas técnicas de *Process Discovery*: Algoritmo Alpha e *Directly Follows Miner*.

2.2.1.1. Algoritmo Alpha

Ao executar o algoritmo Alpha, o resultado a ser obtido é uma *Petri net* que seja capaz de representar todos os *traces* do *eventlog*. Ele faz isso através da análise da ordem de cada uma das atividades em todos os *traces* do *eventlog*, classificando-as todos os pares de atividade de tal maneira (Van der Aalst, 2016):

- Atividades consecutivas restritas (símbolo $>$): uma atividade X é sempre seguida por outra atividade Y;

- Atividades consecutivas opcionais (símbolo \rightarrow): às vezes uma atividade X é seguida por uma outra Y;
- Atividades desconexas (símbolo #): uma atividade X nunca está ligada a uma atividade Y, nem vice-versa;
- Atividades concorrentes (símbolo ||): X é seguida de Y e Y é seguida de X.

Em posse da classificação da relação entre todos os pares de atividades, a *Petri Net* pode ser reconstruída utilizando os padrões encontrados, conforme pode ser visualizado na Figura 4.

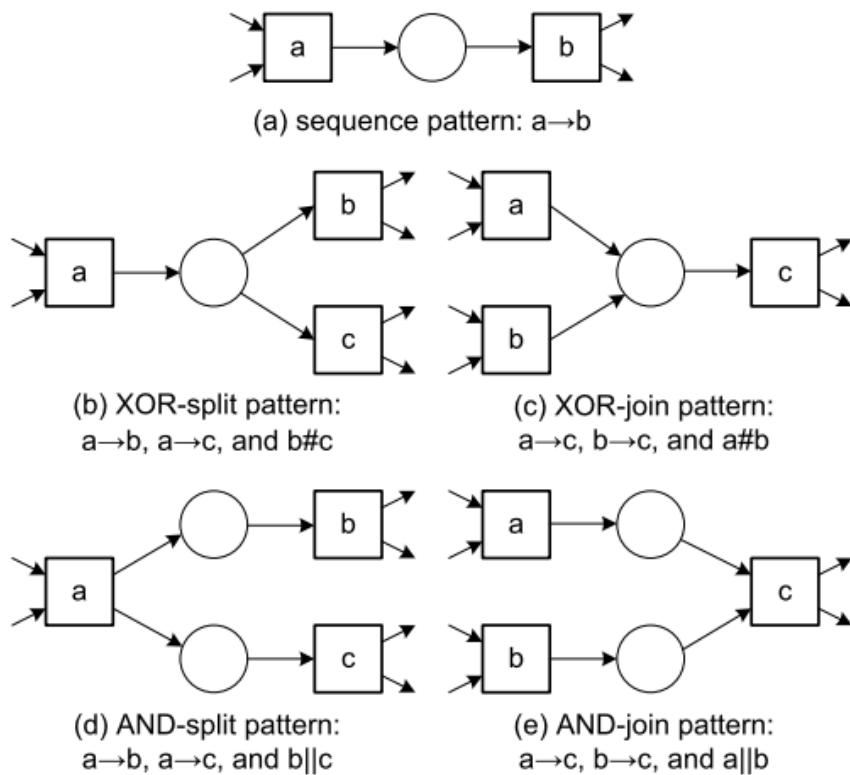


Figura 4 - Padrões típicos do Algoritmo Alpha e sua representação em uma *Petri Net*. Fonte: Van der Aalst (2016).

O algoritmo Alpha é um dos mais simples de Process Discovery, e teve muitas de suas ideias incluídas em outros algoritmos mais avançados. Porém, ele tem dificuldade de representar processos complexos, de lidar com comportamentos que são infrequentes (*noise*) e de compreender *eventlogs* incompletos (muitos poucos *traces*) (Van der Aalst, 2016).

2.2.1.2. Directly Follows Miner

O *Directly Follows Miner* é uma técnica que constrói modelos *Directly Follows* a partir de *eventlogs* considerando exclusivamente a ordem em que as atividades aparecem nos *traces*. Como esse tipo de modelo é incapaz de lidar com concorrência de atividades, um passo essencial em sua construção é remover o *noise*, para que não sejam gerados modelos que não atendam a especificação de *soundness* (Leemans; Poppe; Wynn, 2019).

O método *trace-based*, definido por Leemans, Poppe e Wynn (2019), propõe uma solução para o preparo dos *traces*. Primeiro, é gerado um modelo *Directly follows*, onde todas as atividades que se conectam são desenhadas como transições no modelo, registrando a frequência de cada transição. Depois, selecionam-se as transições menos frequentes e removem-se do *eventlog* todas os *traces* que as contêm, repetindo isso até que atinja um limite máximo pré-definido de *traces* a serem removidos. Assim, o modelo resultante representa uma quantidade aceitável de *traces* do *eventlog*, mantendo ainda sua condição de *soundness* necessária.

2.2.2. Métodos de Avaliação de Modelos

O modelo criado pelos métodos de Process Discovery ou desenhado por especialistas pode ser testado em relação ao *eventlog*, normalmente utilizando o conceito de *play-in*, ou seja, reproduzir os *traces* do *eventlog* seguindo as atividades e transições descritas no modelo (Carmona; Van Dongen; Weidlich, 2022).

Existem quatro métodos de avaliar a efetividade do modelo, descritas a seguir:

- *Fitness*, que é a medida da capacidade do modelo de reproduzir a realidade;
- *Precision*, que é a medida da capacidade do modelo de se restringir a *traces* existentes no *eventlog*;
- *Generalization*, que é a capacidade do modelo de generalizar comportamentos do *eventlog*;
- *Simplicity*, que é a medida de o qual complexo o modelo é (Van der Aalst, 2016).

Neste trabalho, se considerou a utilização do *fitness* para realizar as análises. Portanto, esta métrica será descrita com mais detalhes a seguir.

Fitness consiste na avaliação de quanto o modelo fielmente representa o comportamento visto no *eventlog*. Ele é calculado como a fração de *traces* que são representados pelo modelo sobre todos os *traces* no *eventlog*. Existem alguns algoritmos diferentes que podem calcular numericamente esse valor, no entanto, o método por *alignment* é o principal utilizado por pesquisadores (Carmona; Van Dongen; Weidlich, 2022).

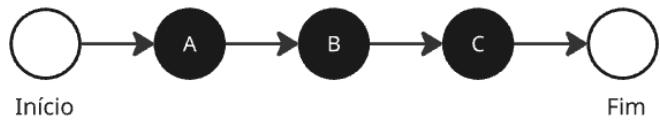
Essa técnica consiste em verificar se o modelo consegue reproduzir cada *trace* do *eventlog* através de uma sequência paralela entre as atividades ocorridas e as possíveis no modelo. Essa correspondência é organizada em uma matriz de duas linhas, onde cada coluna representa um ‘passo’ conjunto: o que ocorreu no log e o que o modelo permitiria naquele ponto (Carmona; Van Dongen; Weidlich, 2022).

Quando uma atividade do log coincide com uma atividade do modelo, ocorre um *synchronous move* (movimento síncrono), sinal de aderência entre execução e especificação. Já quando o modelo prevê uma atividade que não aparece no *trace*, registra-se um *model move* (movimento do modelo), mostrando que algo que deveria ter ocorrido foi omitido. Por outro lado, um *log move* (movimento do log) é quando o log registra a execução de uma atividade que o modelo não autorizaria, sendo então uma ação inesperada (Carmona; Van Dongen; Weidlich, 2022).

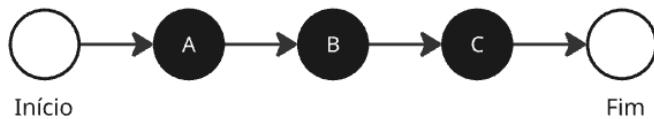
Uma maneira de computar isso numericamente de um único trace (Equação (1)), conforme utilizado pelo *Inductive Visual Miner* (Leemans et al., 2020), é exemplificado pela Figura 5:

$$fitness = \frac{synchronous\ moves}{log\ moves + model\ moves + synchronous\ moves} \quad (1)$$

Modelo de Referência

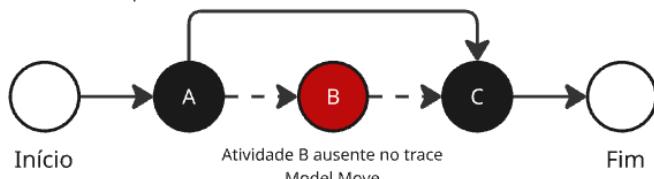


Trace com Fitness = 1



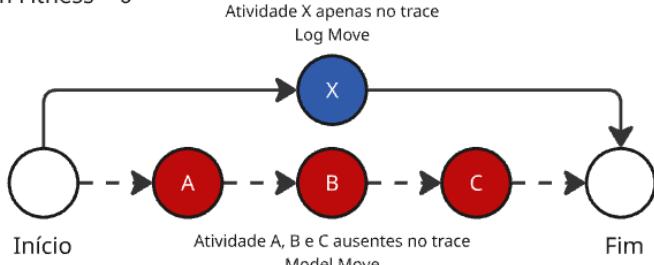
$$\text{fitness} = \frac{3}{0 + 0 + 3}$$

Trace com Fitness = 0,67



$$\text{fitness} = \frac{2}{0 + 1 + 2}$$

Trace com Fitness = 0



$$\text{fitness} = \frac{0}{1 + 3 + 0}$$

Figura 5 - Exemplos de *Fitness*. Fonte: Elaborado pela Autora.

Porém, uma limitação dessa fórmula é a falta de normalização dos denominadores, já que cada trace pode ter uma quantidade diferente de eventos, logo, não há padronização dos valores entre um mesmo *eventlog*, já que cada trace só é comparado em nível individual.

2.3. Ciclo de Vida de Ciência de Dados

Um trabalho de Mineração de Processos pode facilmente fazer uso do Ciclo de Vida de um projeto de ciência de dados. Essa ferramenta é composta por uma sequência de etapas iterativas e fluídas, ligando conhecimentos e desafiando o usuário a aprofundar-

se no projeto. A referência de ciclo de vida para esse projeto escolhida é a de Zumel e Mount (2014), a qual traz essa visão cílica e objetiva para um trabalho de ciência de dados (Figura 6).

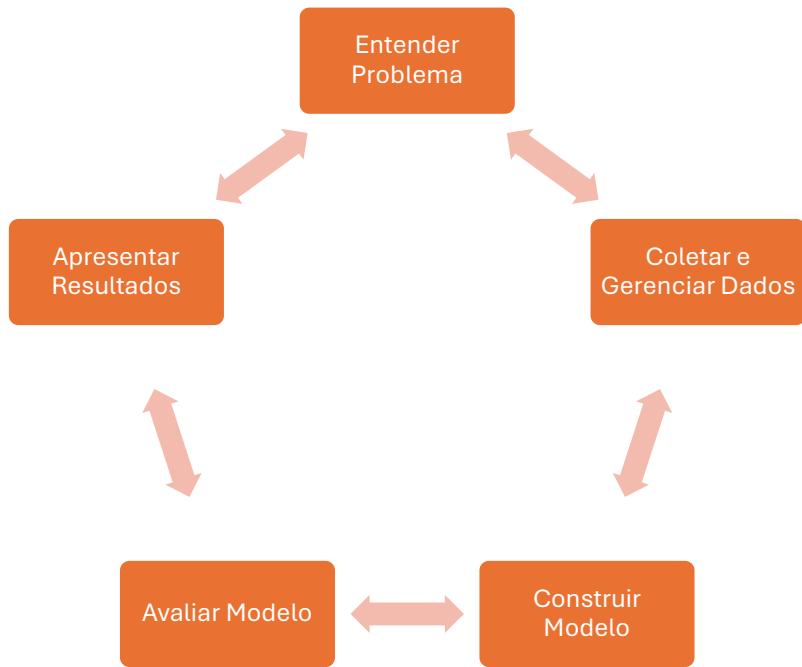


Figura 6 - Ciclo de Vida de Ciéncia de Dados. Adaptado de Zumel e Mount (2014).

O primeiro passo é o Entendimento do Problema: definindo os objetivos centrais e perguntas a serem respondidas durante o trabalho. Em seguida, é necessário compreender quais são os dados necessários para atender aos objetivos do trabalho e obtê-los; essa é a etapa de Coletar e Gerenciar Dados. Dentro dela, também estão a exploração dos dados e o tratamento deles, o que pode trazer novas perguntas e necessidade de novos entendimentos, categorizando a relação iterativa do ciclo de vida do projeto.

O terceiro passo, Construir Modelo, se refere ao processo de extrair insights relevantes utilizando quaisquer técnicas de ciéncia de dados. Dessa forma, progredindo ao próximo passo de Avaliar e Criticar o Modelo, entendendo se o modelo utilizado atende as necessidades do projeto. Novamente, assim como nas outras etapas, podem ser necessárias revisões em dados ou conceitos, para que se possa atingir os objetivos.

Por fim, o ciclo chega à etapa de Apresentação dos Resultados, na qual são expostas as respostas às perguntas definidas na etapa de Entendimento do Problema,

assim como os insights mais interessantes adquiridos durante o desenvolvimento do projeto.

2.4. Departamento de Emergência

No Brasil, um departamento de emergência é entendido como qualquer local onde são realizados Serviços de Atenção às Urgências não Hospitalares. Esses locais têm como objetivo fornecer atendimento a pacientes com condições clínicas graves e não graves, conduzindo avaliação diagnóstica em um primeiro contato para determinar conduta adequada (Brasil, 2025). Para seu funcionamento efetivo, existem requisitos essenciais, como a existência de um fluxo mínimo pré-estabelecido (Figura 7), no qual o paciente deve passar por uma classificação de risco e obrigatoriamente ser avaliado por um médico antes de sua liberação, alta ou internação (Conselho Federal de Medicina, 2014).

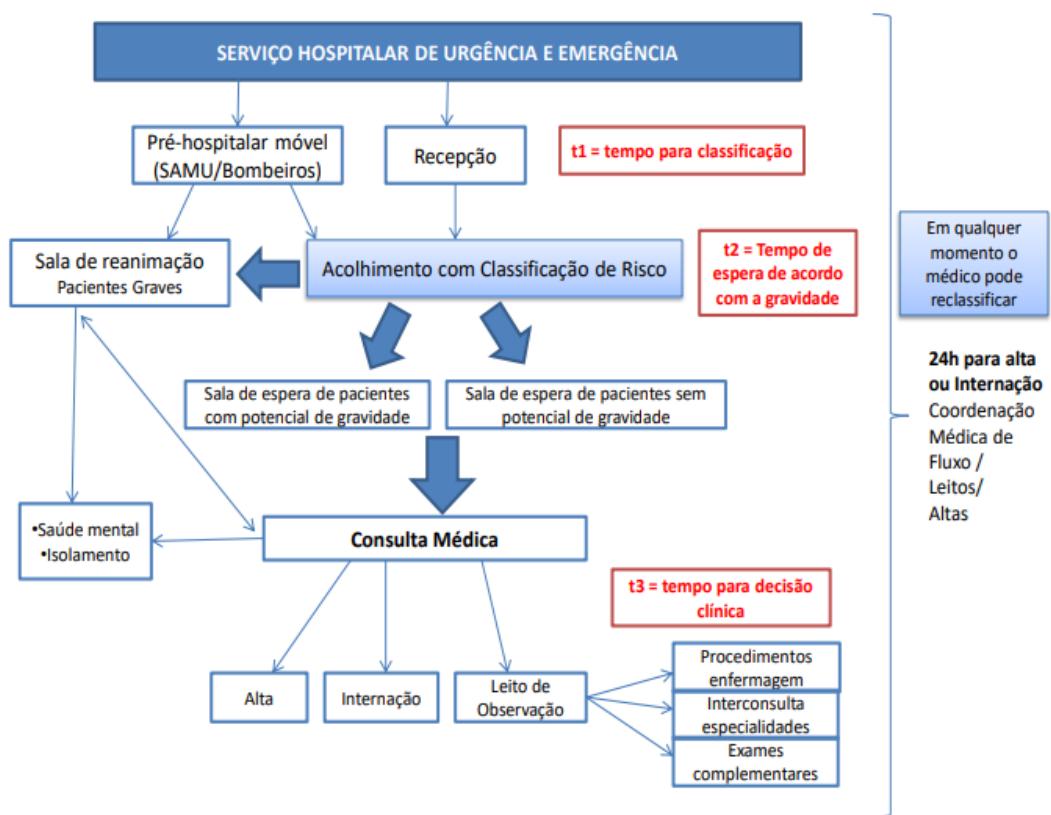


Figura 7 - Algoritmo de fluxo do paciente no Serviço Hospitalar de Urgência e Emergência. Fonte: D'Avila e Batista e Silva (2014)

Os pacientes chegam ao DE por meio de duas fontes principais de entrada dos pacientes, ambulância ou demanda espontânea. Eles são recebidos e, então deve ser feita

uma avaliação do risco (Conselho Federal de Medicina, 2014). Esse processo permite identificar imediatamente quais pacientes necessitam de atendimento urgente e quais podem aguardar, considerando o risco de agravamento do quadro clínico. O método de triagem utilizado por muitos hospitais é o sistema de Manchester um sistema com o objetivo de trazer consenso nos padrões para triagem de pacientes (Manchester Triage Group, 2023). Ele utiliza diversos critérios, como frequência de batimento, oxigenação e histórico, para avaliar a condição do paciente e recomenda o tempo máximo até o primeiro atendimento médico necessário com cinco categorias, como descrito na Tabela 1.

Tabela 1 - Sistema Manchester de Triagem

Número	Nome	Cor	Tempo máximo ¹
1	Imediato	Vermelho	0
2	Muito Urgente	Laranja	10
3	Urgente	Amarelo	60
4	Padrão	Verde	120
5	Não Urgente	Azul	240

1: tempo máximo em minutos para primeiro atendimento médico

Após a classificação de gravidade, o fluxo delimitado pela Figura 7 indica que a próxima atividade deve ser a consulta médica, na qual o paciente será avaliado por um profissional qualificado e será determinado quais serão os próximos passos no tratamento ou diagnóstico. O paciente pode ser internado, receber alta ou continuar no DE recebendo algum tipo de tratamento ou realizando exames complementares, dependendo do resultado da consulta e das decisões médicas. (Conselho Federal de Medicina, 2014).

3. Materiais e Métodos

Nesse trabalho, a metodologia seguida foi o ciclo de vida de ciência de dados, adaptado para um projeto de Mineração de Processos.

3.1. Entendimento do Problema

O primeiro passo, é o entendimento do problema. O trabalho tem como objetivo analisar o fluxo real de um DE hospitalar, portanto, o entendimento de como o processo foi desenhado pelo hospital em questão é o foco dessa etapa inicial.

O hospital parceiro forneceu como insumo para a pesquisa informações sobre o fluxo previsto, representado pela Figura 8.

O paciente primeiro pega uma senha (“Recepção Senha”), momento no qual ele também responde a um questionário para ter uma primeira avaliação de risco conforme o sistema de Manchester. Depois, ocorre a etapa de triagem (“Triagem”), a qual é uma classificação mais assertiva de risco efetuada pelo enfermeiro, comparado à categoria iniciação. Caso o estado do paciente seja crítico (vermelho ou laranja), ele segue para a fila priorizada de atendimento médico, caso contrário, ele tem então suas informações pessoais coletadas ou validadas na recepção (“Recepção Cadastro”) antes de seguir para o atendimento médico (“Atendimento Médico”).

Durante o atendimento, o médico dará uma classificação final de risco conforme sistema de Manchester que será considerada como a mais assertiva na classificação da gravidade. Nesse momento o médico também decidirá se o paciente pode ser liberado (“Alta”) ou se ele precisa de mais ações como tomar medicamentos, fazer exames ou ficar em observação. O paciente que não for liberado, entra nesse ‘loop’ de atividades que podem ser executadas em qualquer ordem: “Exames”, “Medicação”, “Observação”. Porém antes do paciente ser liberado, ele deve retornar ao atendimento médico (“Reavaliação Médica”) para receber um parecer e retornar ao ‘loop’ ou ser liberado/internado (“Alta/Internação”).

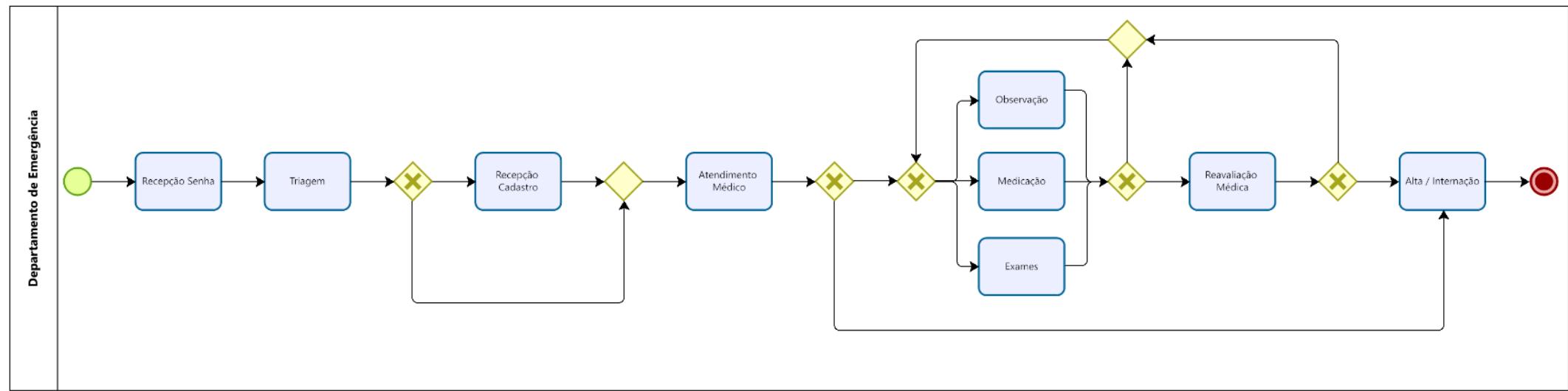


Figura 8 - Modelo BPMN do Fluxo de Pacientes no Departamento de Emergência do Hospital em Questão. Fonte: Elaborado pela Autora.

3.2. Coletar e Gerenciar Dados

Os dados necessários para as análises são o *eventlog* do sistema de informações do hospital e os dados descritivo dos pacientes. Por isso, foram extraídas duas bases de dados abrangendo o período de 01-08-2017 até 31-08-2018 de um DE: a base de dados com as características descritivas dos pacientes e a base de dados contendo o *eventlog*.

A base de dados descritiva continha informações de cada paciente que esteve no DE durante o período (84.837), identificados por um código anonimizado único, contendo atributos de idade, gênero, Código Internacional de Doenças (CID), data de entrada, número e cor dado na triagem e desfecho do caso (alta, óbito, transferência, internação). A base com o *eventlog* continha dados brutos das atividades executadas (1.200.044 registros), contendo o tempo de início e fim de cada atividade, código identificador do paciente e recurso médico utilizado.

A preparação e exploração dos dados foi realizada com o software R Studio versão 4.3.2 (R Core Team, 2023) e com os pacotes *dplyr* (Wickham et al., 2014), *tidyverse* (Wickham, 2016), *readxl* (Wickham; Bryan, 2015), *lubridate* (Spinu; Grolemund; Wickham, 2010).

O primeiro passo foi analisar as atividades previstas no *eventlog* para entender o que era de fato significativo para a análise que esse trabalho propõe. As atividades, em sua maioria tinham cinco estados: “AGUARDANDO ATIVIDADE X”, “CHAMADO ATIVIDADE X”, “EM ATIVIDADE X”, “PAUSADO ATIVIDADE X”, "FINALIZADO ATIVIDADE X". Para simplificar a mineração, foram removidas do *eventlog* os registros de espera, chamado, pausa e finalização das atividades, pois se referem a filas ou início e fim do processo para uso do sistema.

A base não tem uma atividade de entrada, que é quando o paciente primeiro retira a senha e entra no sistema, porém, ele tem uma atividade de “AGUARDANDO ESTRATIFICAÇÃO”, ou seja, aguardando a triagem. Como essa é a primeira atividade esperada no fluxo desenhado pelo hospital, essa atividade foi traduzida em “ADQUIRIU SENHA”. Numa linha parecida, a atividade “EM ESTRATIFICAÇÃO” foi renomeada para ser “TRIAGEM”, de maneira a ficar mais de acordo com o fluxo desenhado.

No *eventlog* existiam várias atividades descrevendo exames de imagem ou laboratoriais: "EM USG", "TOMOGRAFIA COMPUTADORIZADA", "EXAMES

"LABORATORIAIS", "ULTRASSONOGRAFIA", "RESSONANCIA MAGNETICA", "PET CT", "EM LABORATORIO", "EM RX", "RAIO X / EXAME", "EXAMES DE LIQUOR", "EXAMES DIVERSOS", "HEMOPERIVADOS", "EM TOMOGRAFIA", "EM RM", "DOPPLER", "MEDICINA NUCLEAR", "BIOPSIA". Novamente, com o intuito de simplificar a mineração e trazer proximidade do sistema ao desenho, essas atividades foram agrupadas como "EXAME". Foram removidas também os registros em que o paciente tinha como atividade "DISPENSADO", "INATIVO" ou "NÃO COMPARECEU", pois compunham menos de 0,05% de todos os eventos e não representam atividades do processo.

Por fim, as atividades "EM RECEPCAO" e "EM VALIDACAO CADASTRO" foram combinadas em uma só: "EM RECEPCAO CADASTRO", a atividade "EM ATENDIMENTO GESSO" foi agrupada com "EM ATENDIMENTO ENFERMAGEM", e a atividade "EM REAVALIACAO MÉDICA" foi agrupada com "EM ATENDIMENTO MÉDICO". Dessa forma, as atividades que constam no event log são: "ALTA HOSPITALAR", "ADQUIRIU SENHA", "TRIAGEM", "RECEPCAO CADASTRO", "EM ATENDIMENTO MÉDICO", "EM OBSERVAÇÃO", "EM ATENDIMENTO ENFERMAGEM", "EXAME".

No quesito da base de dados descritiva, o tratamento de dados foi mais simples. O primeiro tratamento, foi a remoção de outliers no quesito de tempo de permanência (LOS – *Length of Stay*). Foram removidos da base todos os pacientes que estivessem com o valor de LOS maior que 1,5 vezes o intervalo interquartil (IQR - *Interquartile Range*, definido pela diferença entre o primeiro quartil e o terceiro quartil), logo, todos registros que tiveram tempo de permanência acima de 910,5 minutos (ou acima de 15,175 horas). Essa limpeza retirou 4.988 pacientes (5,9% da base), resultando em uma base de dados com 79.849 pacientes com 566.563 registros no *eventlog*.

Baseado no código internacional de doenças (CID) de cada paciente, foi criada uma coluna com o código do grupo de diagnósticos correspondente (World Health Organization, 2022). Duas novas colunas foram criadas a partir da data de entrada: uma representando o dia da semana e outra com o bloco de horário que o paciente chegou. As horas ficaram divididas com base no plantão dos médicos, vide Tabela 2.

Tabela 2 - Horários de Plantão da Equipe Médica

Período	Turno
Entre 08:00 e 19:59	Dia
Entre 20:00 e 07:59	Noite

No hospital em questão, a quantidade de médicos no DE por turno e dia da semana é conforme descrito na Tabela 3.

Tabela 3 - Quantidade de Médicos por Turno e Dia da Semana no Departamento de Emergência

Médicos por turno	Dia	Noite
Segunda – Quinta	5	4
Sexta – Domingo	3	2

Para a questão da exploração dos dados, e padrões que podem ser identificados, os resultados estão dispostos na Apresentação dos Resultados. Foram feitos gráficos diversos para visualizar o comportamento dos dados por meio do *ggplot2* (Wickham et al., 2007) e *patchwork* (Pedersen, 2019), e realizada uma análise descritiva dos dados, explorando médias, e distribuição das características dos pacientes.

3.3. Construir Modelo e Avaliar Modelo

A próxima etapa no ciclo de vida de ciência de dados consiste na construção do modelo e a seguinte na avaliação do Modelo. Para isso foi utilizado o software de Mineração de Processos ProM Tools versão 6.14 (Van der Aalst et al., 2009) com o plug-in *Inductive Visual Miner* (IVM) (Leemans et al., 2020). Com o IVM foi possível traduzir o modelo desenhado pelo hospital em um modelo *Directly Follows* capaz de ser lido pelo programa, conforme na Figura 9.

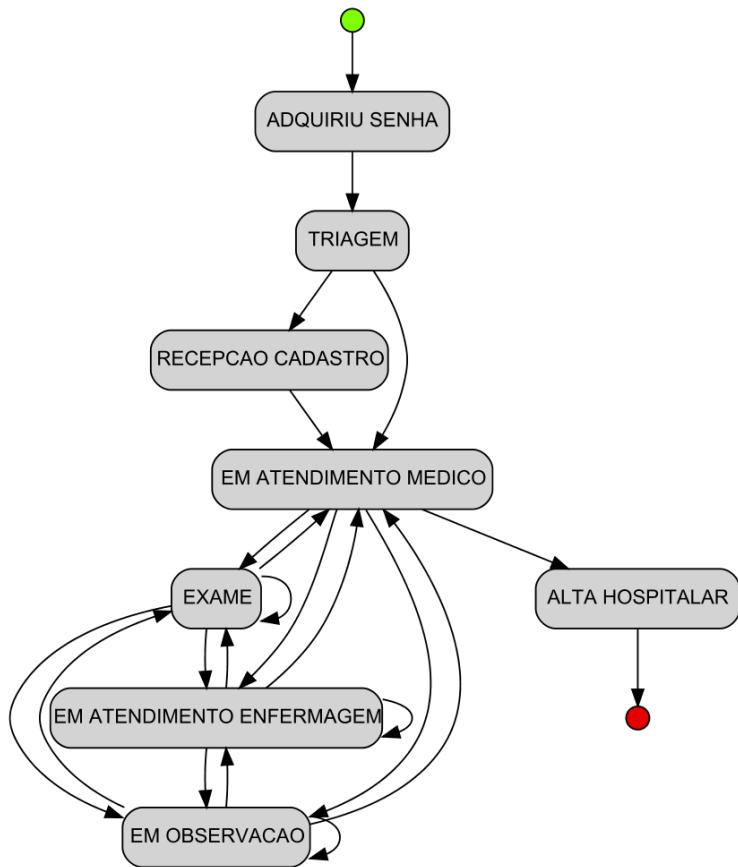


Figura 9 - Modelo *Directly Follows* do Processo do Departamento de Emergência. Fonte: Elaborado pela Autora

Esse modelo então foi utilizado pelo plug-in para fazer os cálculos de *fitness* de cada trace. Com base nisso, foi possível separar os pacientes em três grupos (Grupo A: *fitness* = 1, Grupo B: *fitness* acima da mediana e Grupo C: *fitness* abaixo da mediana) com o intuito de entender as diferenças e semelhanças entre as características dos pacientes por nível de aderência ao fluxo desenhado. Foi montada uma tabela descritiva, semelhante à da etapa de exploração dos dados, comparando as diferenças entre grupos por meio de testes qui-quadrado (F.R.S., 1900), para as variáveis categóricas, e testes Kruskal-Wallis (Kruskal; Wallis, 1952) para as variáveis numéricas.

Para uma próxima etapa de análise, o IVM foi utilizado com bases parciais para minerar modelos *Directly Follows* com os parâmetros de que o processo considerasse todas as atividades (activities = 1) e que filtrasse pelos *traces* 80% mais frequentes (paths = 0,80). Dessa forma, foi possível obter os modelos baseado nos *traces* mais comuns entre os grupos A, B, e C, bem como os modelos por categoria de gravidade.

A última etapa no ciclo de vida de um projeto de Ciência de Dados relevante para esse trabalho é a Apresentação dos Resultados, passo que será exibido na próxima seção.

4. Resultados e Discussão

Nesse Capítulo serão apresentados os resultados obtidos durante as análises, assim como serão levantadas as discussões da significância desses resultados.

4.1. Análise Descritiva

A Tabela 4 mostra, de maneira descritiva, as informações dos pacientes que frequentaram o DE durante o período de 01-08-2017 até 31-08-2018.

Tabela 4 - Tabela Descritiva dos Pacientes

Característica	Pacientes (total = 79.849)
Gênero, n [%]	
Feminino (F)	47.338 [59,3%]
Masculino (M)	32.505 [40,7%]
Indefinido	6 [0,01%]
Idade, n [%]	
Média (σ)	41, (16,75)
Mediana (1Q, 3Q)	38 (29, 52)
Desfecho, n [%]	
Alta	77.209 [96,7%]
Óbito	32 [0,04%]
Internação	1028 [1,29%]
Transferência	1580 [1,98%]
LOS¹ (em minutos)	
Media, (σ)	215, (299)
Mediana (1Q,3Q)	121 (41, 315)
Nível de Gravidade, n [%]	
Vermelho	35 [0,04%]
Laranja	6.784 [8,50%]
Amarelo	27.372 [34,28%]
Verde	37.025 [46,37%]
Azul	8.633 [10,81%]

1: Length of Stay (tempo de estadia na emergência)

Dos quase 80.000 pacientes, pode-se notar que a maior fração (59,3%) deles correspondem ao sexo feminino. É possível também avaliar que a taxa de complicações no desfecho, é muito baixa: ao unir óbitos, internações e transferências, a porcentagem é menor do que 4%. Em relação à gravidade dos pacientes, medida pelo sistema de Manchester, a maior parte dos pacientes (91,46%) se encontra nas gravidades intermediárias e leves, Amarelo, Verde e Azul, com uma menor parte (8,54%) compondo os grupos mais graves, Vermelho e Laranja.

Em relação às idades dos pacientes, foi analisado por meio de *boxplots* a distribuição de idade de acordo com a classificação de gravidade. Os resultados podem ser observados na Figura 10.

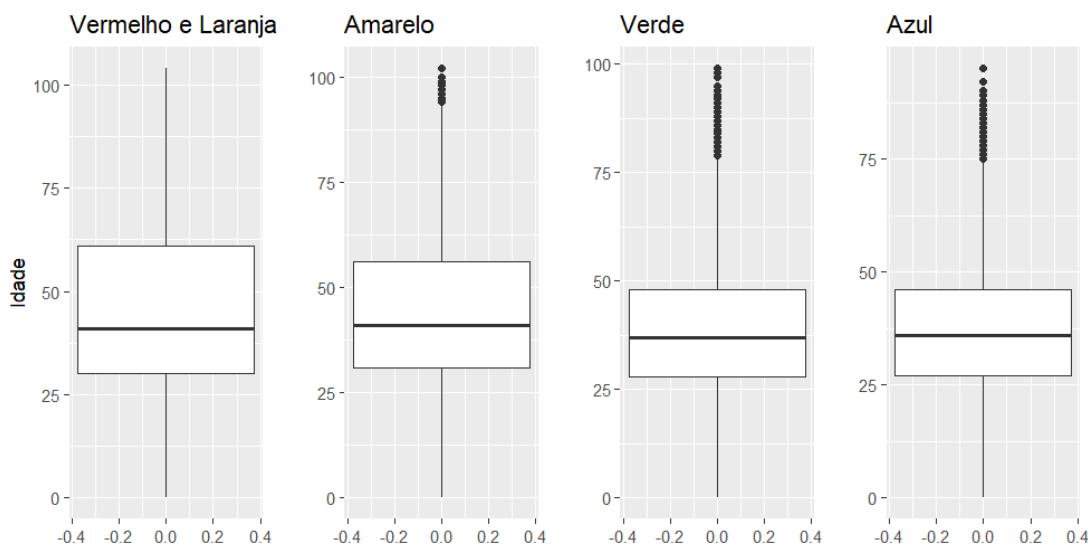


Figura 10 - *Boxplot* de Distribuição de Idades por Categoria de Gravidade. Fonte: Elaborado pela Autora.

A diferença de idade entre os grupos é estatisticamente significativa, conforme indicado pelo teste de Kruskal-Wallis ($p < 0,01$). Observa-se também as medianas são diferentes entre os grupos, mostrando um aspecto agravante das idades mais avançadas, além de existir uma diferença na média: para a categoria Vermelho a média é de 53,0 anos; para Laranja, 45,9 anos; Amarelo, 44,4 anos; Verde, 39,3 anos; e Azul, 37,7 anos. Além disso, nas categorias Vermelho e Laranja existe variação maior das idades, evidenciado pelo maior IQR, mostrando que há uma grande variabilidade. Nas categorias

Amarelo, Verde e Azul, o formato do *boxplot* está mais concentrado, com idades menos variáveis.

Outro aspecto importante a destacar é a frequência de chegada dos pacientes ao longo do período em questão. A Figura 11 mostra a quantidade de pacientes por mês analisado no DE, que apresentou um leve crescimento a partir do final de 2017.

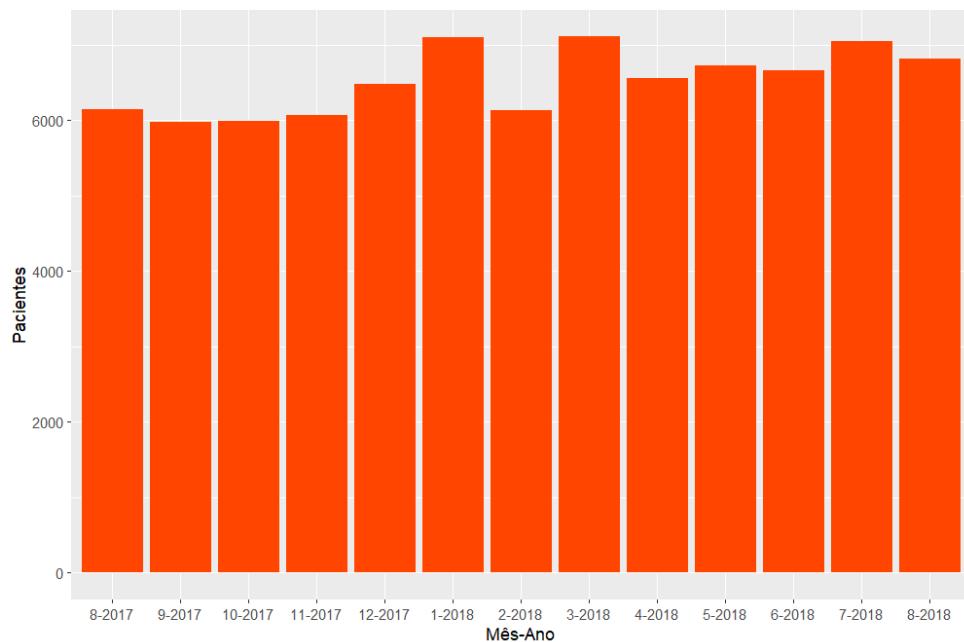


Figura 11 – Volume Acumulado de Pacientes no Departamento de Emergência por Mês e Ano. Fonte: Elaborado pela Autora.

Por outro lado, a Figura 12 mostra o número das entradas durante os dias da semana, começando com um maior volume na segunda-feira, e diminuindo progressivamente até domingo.

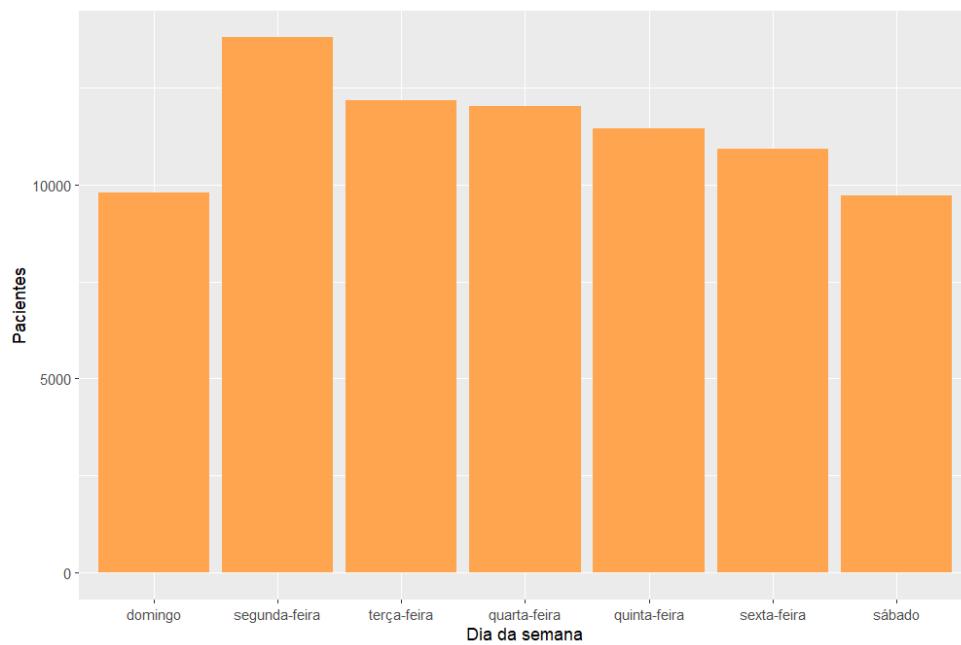


Figura 12 – Volume Acumulado de Pacientes no Departamento de Emergência por Dia da Semana. Fonte: Elaborado pela Autora.

Na Figura 13, é exibido o valor acumulado de pacientes chegando no DE em cada hora no dia, mostrando que os horários mais frequentes de chegada é durante o turno diurno (8:00 às 19:59). Informação complementada pela Figura 14, a qual mostra a quantidade de pacientes em cada dia da semana e turno. A quantidade de pacientes que chegam durante o turno do dia é consideravelmente maior (cerca de 3 vezes maior) comparado ao turno da noite em todos os dias da semana.

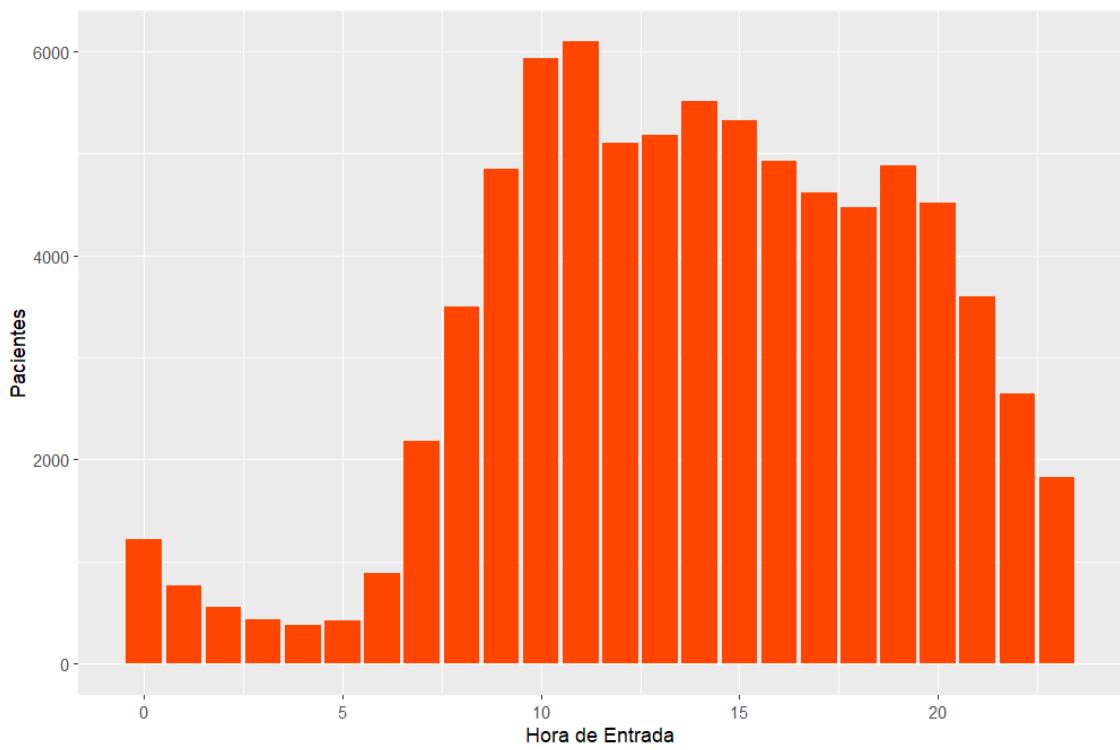


Figura 13 – Volume Acumulado de Pacientes no Departamento de Emergência por Hora de Entrada.
Fonte: Elaborado pela Autora.

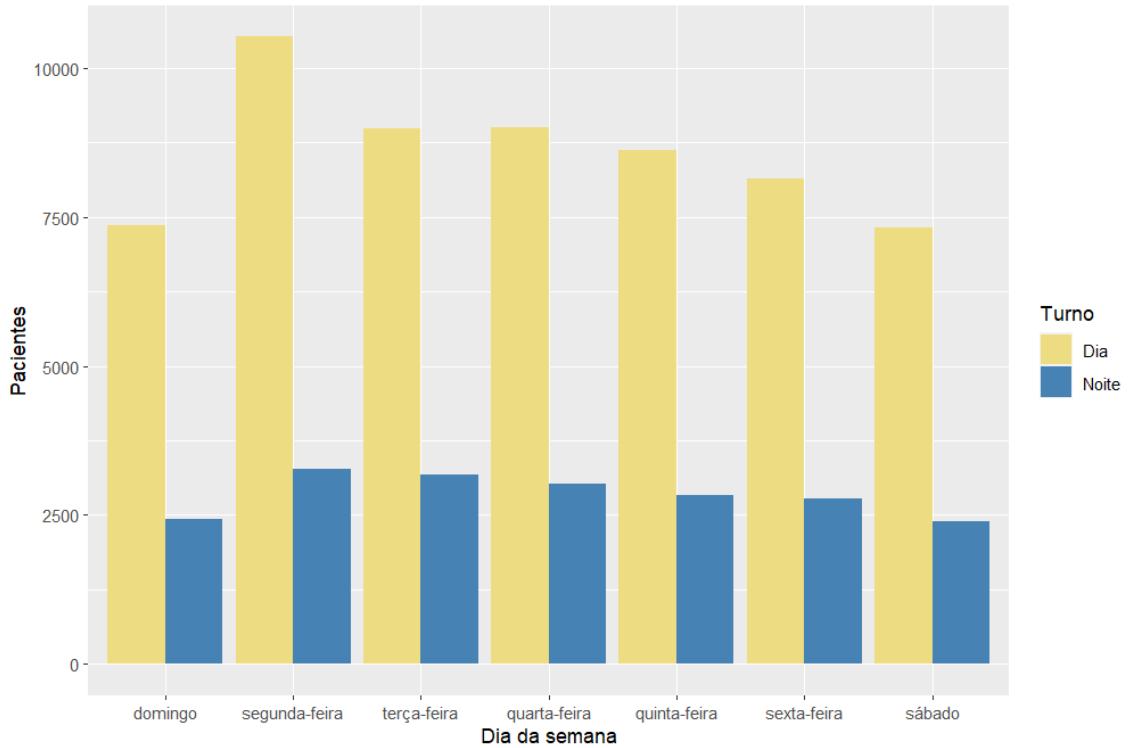


Figura 14 – Volume Acumulado de Pacientes no Departamento de Emergência por Turnos de Entrada e Dia da Semana. Fonte: Elaborado pela Autora.

Por fim, a Figura 15 mostra que a porcentagem mais significativa, média de 14,4%, dos pacientes com categorias mais graves (Vermelho e Laranja) ocorre nos horários da madrugada (entre meia noite e 5 horas), comparado ao resto dos horários na qual a porcentagem média é 8,2%.

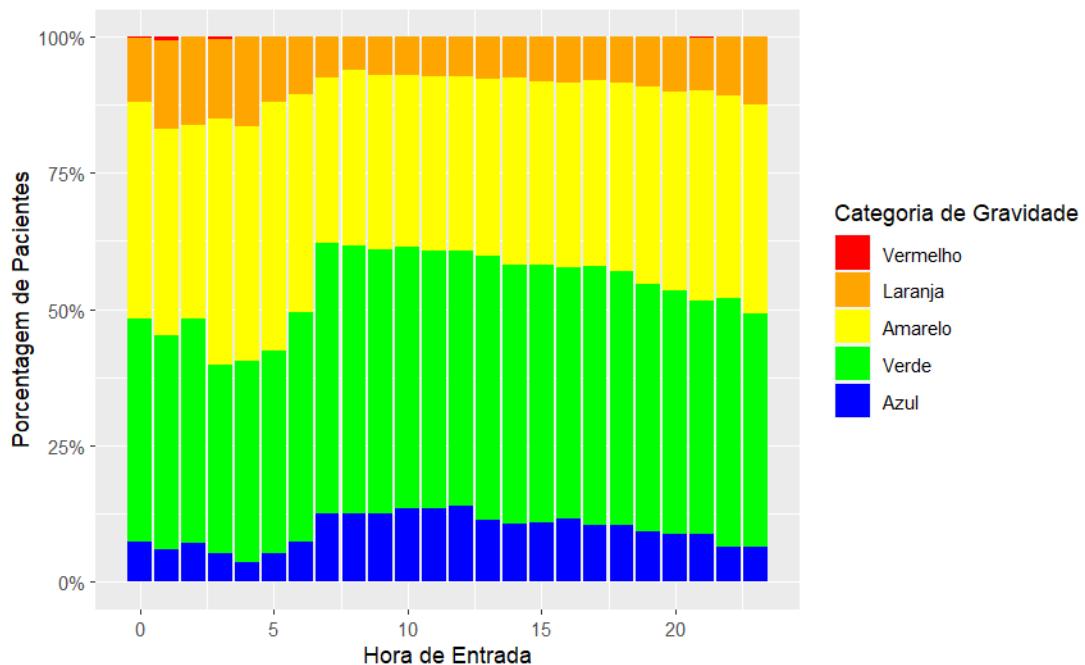


Figura 15 – Porcentagem de Níveis de Gravidades no Departamento de Emergência por Hora de Entrada.
Fonte: Elaborado pela Autora.

4.2. Análise de Conformidade

Ao utilizar o plug-in IVM do ProM Tools (Leemans et al., 2020) para avaliar a aderência dos *traces* dos pacientes em relação ao modelo estabelecido pelo hospital, foi possível obter a distribuição de *fitness* do modelo, apresentada na Figura 16. A distribuição dos *traces* tem cerca de 1/3 dos casos classificados como perfeitamente aderentes ao fluxo, sem desvios; o resto se encontra distribuído entre 0,4 e 0,95, com uma grande concentração dos casos com mais de 0,80 de *fitness*. Esses dados trazem a informação de que a maior parte dos casos que tem desvios ($fitness < 1$) não tem grandes diferenças em relação ao fluxo desenhado, devido ao valor alto de *fitness* mantido (mediana 0,86).

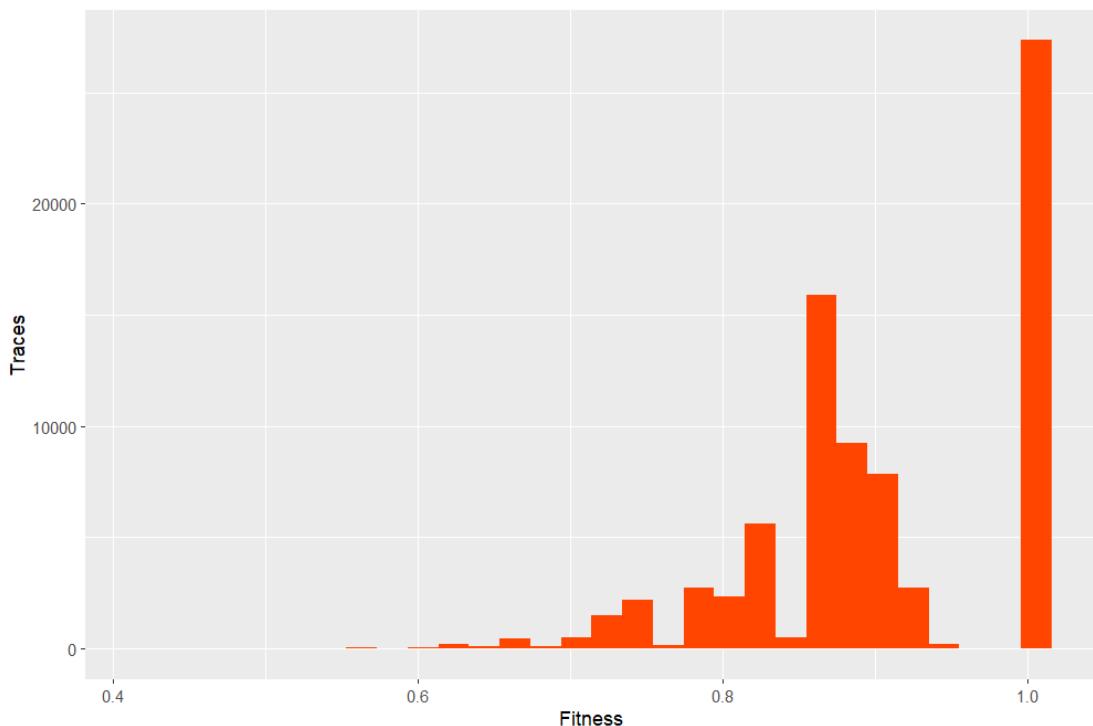


Figura 16 - *Fitness* do Trace de cada Paciente. Fonte: Elaborado pela Autora.

A Tabela 5 mostra a análise descritiva dos pacientes em relação à aderência ao modelo, dividido nos seguintes grupos: os com aderência perfeita (*fitness* = 1), e o restante dos pacientes separados pela mediana do *fitness* (0,86).

Tabela 5 - Tabela Descritiva por Fitness

Características	Grupo A: <i>Fitness</i> = 1 (n= 27.359)	Grupo B <i>Fitness</i> ≤ 0,86 (n= 26.062)	Grupo C <i>Fitness</i> < 0,86 (n= 26.428)	<i>p</i> ¹
Gênero, n [%]				
Feminino (F)	16.157 [59,0%]	15.580 [59,7%]	15.601 [59,0%]	0,412
Masculino (M)	11.200 [40,9%]	10.480 [40,2%]	10.825 [40,9%]	
Indefinido	2 [0,01%]	2 [0,01%]	2 [0,01%]	
Idade				<0,01
Média (σ)	38,7 (14,9)	43,9 (17,5)	41,8 (17,3)	
Mediana (1Q, 3Q)	37 (28,47)	41 (31,55)	38 (29, 52)	
Desfecho, n [%]				<0,01

Alta	27.351 [99,9%]	24.731 [94,9%]	25.127 [95,1%]	
Óbito	7 [0,02%]	8 [0,03%]	17 [0,06%]	
Internação	0	513 [1,97%]	515 [1,95%]	
Transferência	1 [0,00%]	810 [3,10%]	769 [2,91%]	
LOS² (minutos)				<0,01
Media (σ)	45,9 (41,2)	322 (227)	285 (246)	
Mediana (1Q,3Q)	35 (22,55)	241 (140,472)	200 (78,448)	
Nível de Gravidade, n [%]				<0,01
Vermelho	0	1 [0,00%]	34 [0,13%]	
Laranja	543 [1,98%]	1.397 [5,36%]	4.844 [18,3%]	
Amarelo	4.034 [14,7%]	17.751 [68,1%]	5.587 [21,1%]	
Verde	16.072 [58,7%]	6.719 [25,9%]	14.234 [53,8%]	
Azul	6.710 [24,5%]	195 [0,75%]	1.729 [6,54%]	
Turno				<0,01
Dia	22.122 [80,8%]	18.803 [72,1%]	19.044 [72,0%]	
Noite	5.237 [19,2%]	7.259 [27,9%]	7.384 [28,0%]	
Dia da Semana				<0,01
Domingo	3.485 [12,7%]	3.130 [12,0%]	3.170 [12,0%]	
Segunda	4.261 [15,6%]	4.690 [18,0%]	4.841 [18,3%]	
Terça	4.354 [15,9%]	3.886 [14,9%]	3.934 [14,9%]	
Quarta	4.690 [17,1%]	3.604 [13,8%]	3.727 [14,1%]	
Quinta	3.879 [14,2%]	3.790 [14,5%]	3.781 [14,3%]	
Sexta	3.335 [12,2%]	3.851 [14,8%]	3.730 [14,1%]	
Sábado	3.355 [12,3%]	3.111 [11,9%]	3.245 [12,3%]	

1: p-valor dos testes qui-quadrado para variáveis categóricas e Kruskal-Wallis para variáveis numéricas

2: Length of Stay (tempo de estadia)

Nessa Tabela, algumas informações ganham destaque. Em relação à idade dos pacientes, aqueles com menor aderência ao fluxo possuem idades com o IQR maior (A: 19, B: 24, C: 23), adicionando variabilidade nos casos. A taxa de complicações no desfecho também é relativamente maior (A: 0,02%, B: 5,10%, C: 4,92%) o que pode explicar o porquê dessa divergência do fluxo padrão.

Ademais, o tempo de permanência no hospital cresce consideravelmente para os casos em que a aderência ao processo não é perfeita, o que pode ter correlação com diversos fatores, como idade e gravidade do caso. Como pode-se ver que a concentração de casos graves está no grupo menos aderente, com as categorias Vermelho e Laranja compondo apenas 1,98% do grupo A e 18,4% do grupo C.

Finalmente, também é relevante ressaltar que, no grupo A, a distribuição de entrada ao longo dos dias da semana é relativamente diferente do que foi visto na Figura 12, com o dia mais cheio sendo a quarta-feira, seguido pela terça-feira, e apenas então a segunda-feira aparece.

4.3. Process Discovery

Com isso, surge a necessidade de entender também em quais etapas esses fluxos não aderentes estão se divergindo do modelo. Isso pode ser feito através da aplicação de *Process Discovery* de cada um dos grupos. Utilizando o IVM e a modelagem *Directly Follows* e considerando apenas os *traces* 80% mais frequentes de cada grupo foi possível obter o desenho do processo executado.

Para o grupo A (Figura 17), o processo minerado é o mais simples possível, explicando porque esses pacientes tinham aderência alta. Apenas um pequeno número de pacientes (2.280 ou 8%) realiza exames, recebe medicamento ou fica em observação, configurando então, um fluxo direto que passa por todas as etapas simples.



Figura 17 - Modelo Minerado do Grupo A. Fonte: Elaborado pela Autora.

O grupo B (Figura 18), no entanto, já apresenta divergências: na etapa seguinte ao atendimento médico, o programa minerou o fluxo com exames, medicamentos e em observação. O motivo desse grupo não estar completamente aderente é o fato de não ter

um registro de retorno ao atendimento médico antes da alta hospitalar. Isso pode se configurar em um erro de dados, que têm as informações após o primeiro atendimento médico nebulosas.

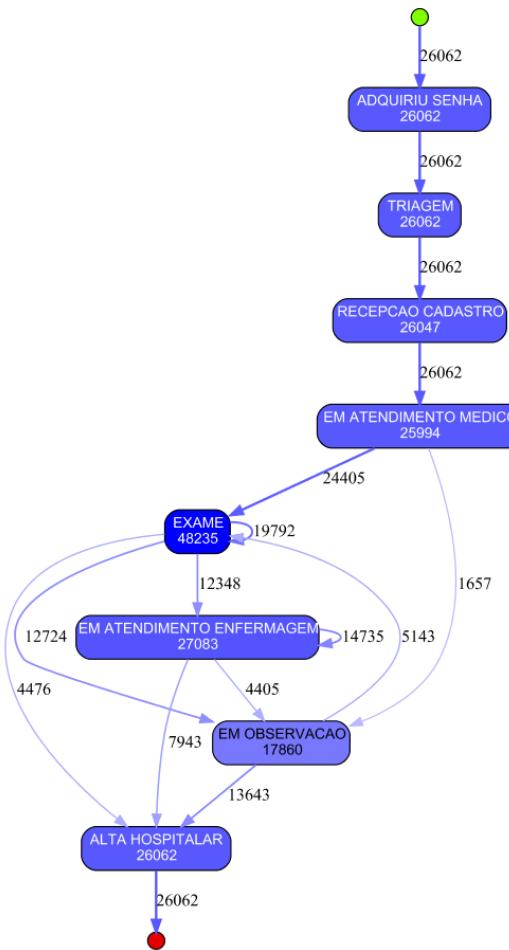


Figura 18 - Modelo Minerado do Grupo B. Fonte: Elaborado pela Autora.

Por fim, para o grupo C (Figura 19) foi minerado um processo semelhante ao do grupo B, porém com ressalvas. A primeira delas é o fato de ter repetição nas atividades de Triagem e Recepção/Cadastro, o que não é esperado em um processo real e pode-se indicar uma inconsistência nos dados. O próximo é o fato de termos certos pacientes que pulam a etapa de atendimento médico e já passam para o loop de tratamento, conforme descrito na Tabela 5.

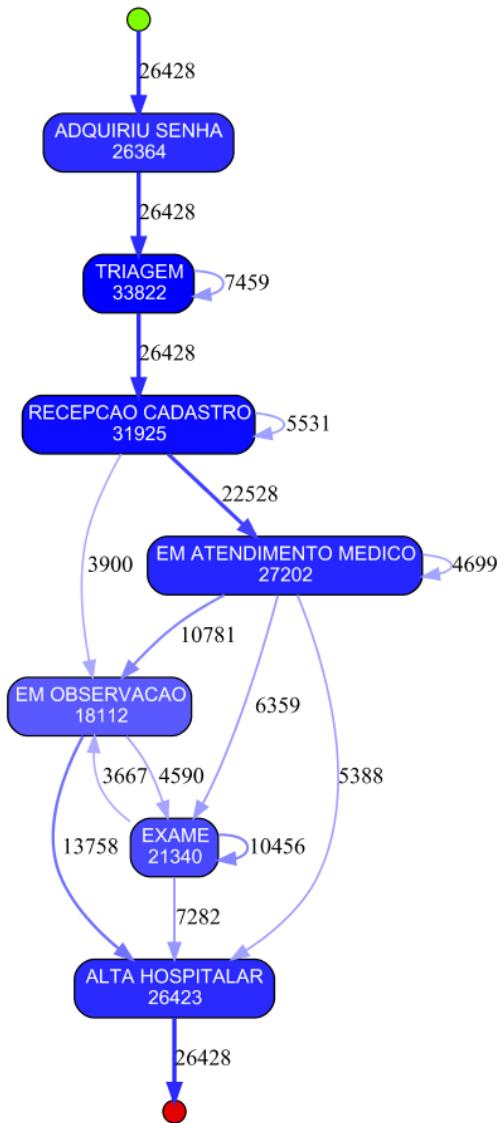


Figura 19 - Modelo Minerado do Grupo C. Fonte: Elaborado pela Autora.

Como a gravidade do paciente aparenta ser um determinante na aderência do fluxo que o paciente segue, foi feito mais uma rodada de análises de Mineração de Processos, dividindo a base de dados pela categorização de gravidade. Começando pela gravidade mais amena, Azul (Figura 20), podemos ver que ela tem o formato igual ao da Figura 17 ($fitness = 1$), com a exceção da repetição da atividade de cadastro, mostrando que o paciente menos grave, possui de fato o fluxo mais simples.



Figura 20 - Modelo Minerado dos Pacientes Categoria Azul. Fonte: Elaborado pela Autora.

As gravidades Verde (Figura 21) e Amarela (Figura 22), são bem semelhantes, com a exceção do fato que mais de 70% dos pacientes Amarelo necessitam de exames após a consulta, contrastando com os 24% da categoria Verde.

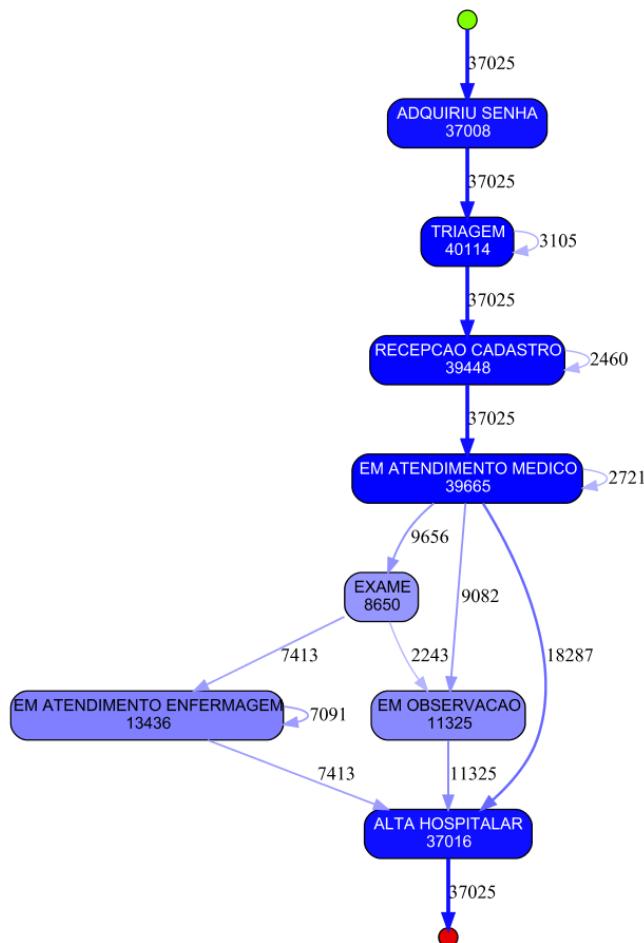


Figura 21 - Modelo Minerado dos Pacientes Categoria Verde. Fonte: Elaborado pela Autora.

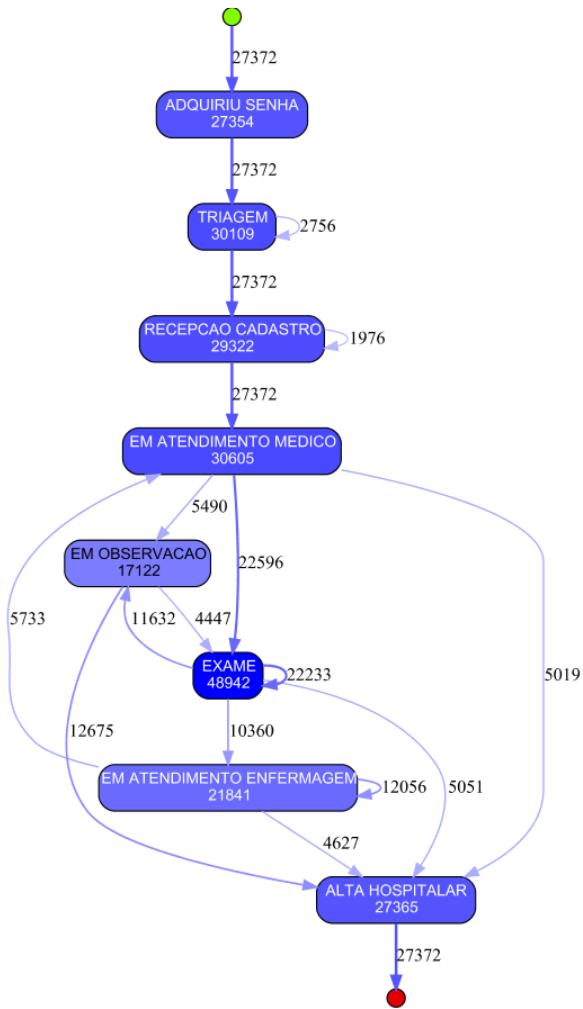


Figura 22 - Modelo Minerado de Pacientes Categoria Amarelo. Fonte: Elaborado pela Autora.

Por fim, a Figura 23 mostra o processo minerado dos pacientes de maior gravidade: Laranjas e Vermelhos. Assim como o grupo C, na Figura 19, a aderência ao fluxo desenhado é baixa, devido a um número parecido de pacientes (3900 e 3700) não realizam a etapa de atendimento médico após a etapa de Recepção Cadastro e seguem direto para a etapa de observação. Pode-se entender que isso ocorre para o grupo C assim como para o grupo das gravidades Laranja e Vermelho devido à alguma necessidade crítica de atendimento imediato.

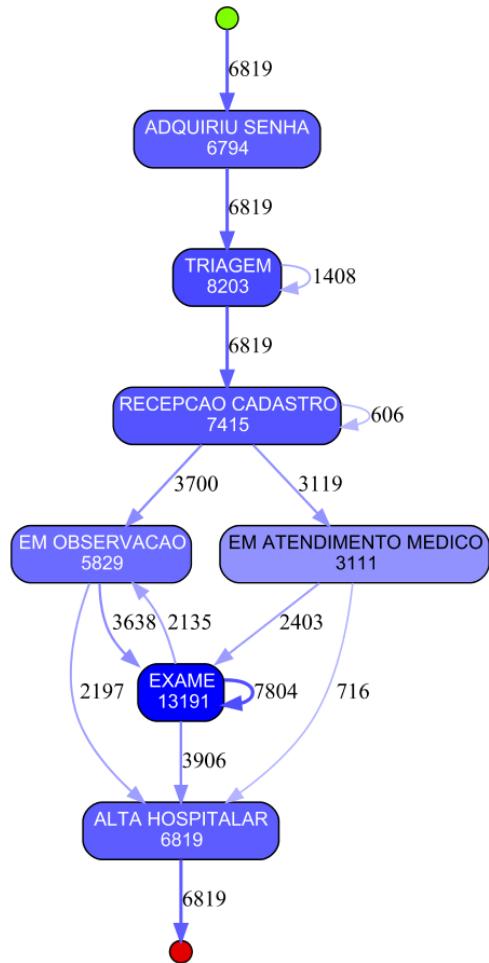


Figura 23 - Modelo Minerado dos Pacientes Categoria Laranja e Vermelho. Fonte: Elaborado pela Autora.

4.4. Considerações Adicionais

Alguns pontos interessantes a se considerar a partir desses resultados são a prevalência de exames como uma etapa fundamental, principalmente para aqueles pacientes com gravidade maior. É importante ressaltar que a atividade de exame corresponde a um grande uso de recursos do hospital, e sendo uma atividade que ocorre com muita frequência é interessante de se pensar em como otimizar a alocação de recursos hospitalares para melhor atender os pacientes.

Ademais, sobre as estatísticas de resolução do hospital, podemos ver que esse é um DE com métricas muito boas, avaliando o tempo entre chegada e o primeiro atendimento médico, pode-se ver quantos porcento dos pacientes são atendidos no tempo recomendado (Manchester Triage Group, 2023) (Tabela 6):

Tabela 6 - Porcentagem de Pacientes Atendidos a Tempo por Gravidade

Nº	Nome	Cor	Tempo máximo ¹	Atendido a tempo ²
1	Imediato	Vermelho	0	Dados insuficientes
2	Muito Urgente	Laranja	10	11,48%
3	Urgente	Amarelo	60	88,28%
4	Padrão	Verde	120	99,52%
5	Não Urgente	Azul	240	91,82%

1: tempo máximo em minutos para primeiro atendimento médico 2: Porcentagem de pacientes Atendido por um médico dentro do tempo máximo recomendado pelo Sistema de Manchester

5. Conclusão

A análise conduzida neste trabalho permitiu compreender o comportamento operacional do Departamento de Emergência (DE) estudado, bem como avaliar a conformidade entre o fluxo processual definido pelo hospital e o fluxo executado pelos pacientes a partir dos dados. O DE atende um volume alto de pacientes, com a maioria apresentando níveis de gravidade mais leves, apresenta baixa taxa de desfechos com óbitos e exibe padrões de demanda, principalmente diurna e no início da semana, além de maior proporção de casos graves durante a madrugada. Esses achados orientam o hospital para um planejamento responsivo de recursos e de escalas de trabalho ajustadas às variações de demanda.

A avaliação do *fitness* dos *traces* mostrou que cerca de um terço dos pacientes segue exatamente o fluxo modelado, enquanto a maior parte dos demais apresenta apenas pequenas divergências, com mediana de *fitness* com desvios de 0,86. No entanto, esses desvios podem ser explicados em grande parte pela característica clínica dos pacientes: em geral, são mais graves, apresentam maior taxa de complicações e permanecem mais tempo no hospital. Ademais, pacientes mais graves tendem a ser mais velhos e apresentar maior variabilidade etária, o que se configura em mais um agravante de condição.

As técnicas de descoberta de processos através do IVM auxiliaram a visualização e modelagem dos processos reais de diferentes perfis de pacientes. Enquanto pacientes menos graves apresentam fluxos simples e diretos, aqueles com maior gravidade seguem caminhos mais complexos, com uso intensivo de recursos. Isso mostra a necessidade de avaliar continuamente a capacidade e os recursos disponíveis para os DE, especialmente considerando os padrões de demanda.

Além disso, foram identificadas inconsistências relevantes no *eventlog*, que podem trazer ruídos para os resultados. Do ponto de vista gerencial, esses resultados destacam a importância de aprimorar a qualidade e a completude dos registros hospitalares, pois essas informações impactam diretamente análises operacionais e tomadas de decisão.

Desse modo, este trabalho mostra que, apesar de o DE em questão possuir um processo bem estruturado e bons resultados de performance, a maior complexidade dos casos graves e algumas falhas de registro afetam a aderência ao modelo processual teórico e revelam pontos de atenção para as operações. Além disso, mostra como a Mineração de

Processos é uma ferramenta útil para entender o fluxo real dos pacientes, identificar desvios importantes, orientar o planejamento de recursos e apoiar estratégias que aumentem a eficiência e a qualidade do atendimento em saúde.

Em relação a estudos futuros, este trabalho abre espaço para algumas pesquisas adicionais. Uma possibilidade é utilizar as técnicas e metodologias aqui aplicadas para abranger outros departamentos de emergência, principalmente no setor público, que pode ter uma limitação de recursos diferente do hospital analisado. Outra possibilidade de pesquisa é o uso de modelos de simulação de eventos discretos ou simulação de agentes para a exploração de diferentes cenários operacionais com base nos processos minerados nesse trabalho, buscando otimizações de alocação de recursos e performance. Por fim, uma análise sistêmica de quais são os motivos para as inconsistências vistas na base seria interessante para garantir melhor qualidade de *eventlogs*, permitindo análises ainda mais refinadas de Mineração de Processos no futuro.

6. Bibliografia

BRASIL. UPA 24h. Disponível em: <<https://www.gov.br/saude/pt-br/assuntos/saude-de-a-a-z/u/upa-24h>>. Acesso em: 20 nov. 2025.

CARMONA, Josep; VAN DONGEN, Boudewijn; WEIDLICH, Matthias. Conformance Checking: Foundations, Milestones and Challenges. In: **Lecture Notes in Business Information Processing**. [S.l.]: Springer Science and Business Media Deutschland GmbH, 2022. v. 448 p. 155–190.

CONSELHO FEDERAL DE MEDICINA. **RESOLUÇÃO CFM nº 2.077/2014**. D.O.U.BrasilConcelho Federal de Medicina, , 16 set. 2014.

DALLAGASSA, Marcelo Rosano *et al.* Opportunities and challenges for applying process mining in healthcare: a systematic mapping study. **Journal of Ambient Intelligence and Humanized Computing**, v. 13, n. 1, p. 165–182, 8 jan. 2022.

DIBA, Kiarash *et al.* Extraction, correlation, and abstraction of event data for process mining. **Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery**, v. 10, n. 3, p. e1346, 1 maio 2020.

DUMAS, Marlon *et al.* **Fundamentals of business process management: Second Edition**. Segunda Edição ed. [S.l.]: Springer Berlin Heidelberg, 2018.

F.R.S., Karl Pearson. X. On the criterion that a given system of deviations from the probable in the case of a correlated system of variables is such that it can be reasonably supposed to have arisen from random sampling. **The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science**, v. 50, n. 302, p. 157–175, jul. 1900.

GARCIA, Cleiton dos Santos *et al.* **Process mining techniques and applications – A systematic mapping study.** *Expert Systems with Applications* Elsevier Ltd, , 1 nov. 2019.

KRUSKAL, William H.; WALLIS, W. Allen. Use of Ranks in One-Criterion Variance Analysis. **Journal of the American Statistical Association**, v. 47, n. 260, p. 583–621, 1952.

LEEMANS, Sander J. J. *et al.* **Visual Miner.** , 5 jun. 2020. Disponível em: <<https://leemans.ch/visualminer/home/>>. Acesso em: 6 nov. 2025

LEEMANS, Sander J. J.; POPPE, Erik; WYNN, Moe T. Directly Follows-Based Process Mining: Exploration & a Case Study. *In:* IEEE, jun. 2019. Disponível em: <<https://ieeexplore.ieee.org/document/8786057/>>. Acesso em: 28 out. 2025

MANCHESTER TRIAGE GROUP. **Emergency Triage.** 3. ed. [S.l.]: John Wiley & Sons, Ltd, 2023.

NEIRA, Ricardo Alfredo Quintano *et al.* Analysis and Optimization of a Sepsis Clinical Pathway Using Process Mining. **Lecture Notes in Business Information Processing**, v. 362 LNBP, p. 459–470, 2019.

OBJECT MANAGEMENT GROUP. Business Process Model and Notation (BPMN). 2010.

ORTÍZ-BARRIOS, Miguel Angel; ALFARO-SAÍZ, Juan José. Methodological Approaches to Support Process Improvement in Emergency Departments: A Systematic

Review. **International Journal of Environmental Research and Public Health 2020, Vol. 17, Page 2664**, v. 17, n. 8, p. 2664, 13 abr. 2020.

PEDERSEN, Thomas Lin. **patchwork: The Composer of Plots. CRAN: Contributed Packages**, 1 dez. 2019.

PEGORARO, Fábio *et al.* Integrando a mineração de processos e simulação para apoiar a gestão de fluxo de pacientes em Pronto-Socorro. **Revista de Gestão e Secretariado (Management and Administrative Professional Review)**, v. 14, n. 4, p. 6260–6274, 27 abr. 2023.

PEREIRA, Gustavo Bernardi; SANTOS, Eduardo Alves Portela; MACENO, Marcell Mariano Corrêa. Process mining project methodology in healthcare: a case study in a tertiary hospital. **Network Modeling Analysis in Health Informatics and Bioinformatics**, v. 9, n. 1, p. 28, 15 dez. 2020.

R CORE TEAM. **R: A language and environment for statistical computing . R Foundation for Statistical Computing** Vienna, Austria, 2023. Disponível em: <<https://www.R-project.org>>. Acesso em: 6 nov. 2025

SPINU, Vitalie; GROLEMUND, Garrett; WICKHAM, Hadley. **lubridate: Make Dealing with Dates a Little Easier. CRAN: Contributed Packages**, 15 ago. 2010.

VAN DER AALST, Wil. **Process mining: Data science in action**. Segunda Edição ed. [S.l.]: Springer Berlin Heidelberg, 2016.

VAN DER AALST, Wil M. P. *et al.* ProM: The Process Mining Toolkit. In: DE MEDEIROS, Ana Karla A.; WEBER, Barbara (orgs.). Ulm, Germany: CEUR, 8 set. 2009.

VAN DER AALST, Wil M. P.; DUSTDAR, Schahram. Process Mining Put into Context. **IEEE Internet Computing**, v. 16, n. 1, p. 82–86, jan. 2012.

WICKHAM, Hadley *et al.* **ggplot2: Create Elegant Data Visualisations Using the Grammar of Graphics.** CRAN: Contributed Packages, 1 jun. 2007.

WICKHAM, Hadley *et al.* **dplyr: A Grammar of Data Manipulation.** CRAN: Contributed Packages, 16 jan. 2014.

WICKHAM, Hadley. **tidyverse: Easily Install and Load the “Tidyverse”.** CRAN: Contributed Packages, 9 set. 2016. Disponível em: <<https://CRAN.R-project.org/package=tidyverse>>

WICKHAM, Hadley; BRYAN, Jennifer. **readxl: Read Excel Files.** CRAN: Contributed Packages, 14 abr. 2015.

WORLD HEALTH ORGANIZATION. **ICD-11: International classification of diseases (11th revision).** Disponível em: <<https://icd.who.int/en/>>. Acesso em: 20 nov. 2025.

ZUMEL, Nina.; MOUNT, John. **Practical data science with R.** [S.l.]: Manning, 2014.