

UNIVERSIDADE FEEVALE

GABRIEL EDUARDO MARTINI

O APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DE STATUS GLICÊMICO DE  
PACIENTES

*Anteprojeto de Trabalho de Conclusão*

Novo Hamburgo  
2019

Gabriel Eduardo Martini

O APRENDIZADO DE MÁQUINA NA CLASSIFICAÇÃO DE STATUS GLICÊMICO DE  
PACIENTES

Anteprojeto de Trabalho de Conclusão de  
curso, apresentado como requisito parcial  
à obtenção do grau de Bacharel em  
Ciência da Computação pela  
Universidade Feevale

Orientador: Prof. Dr. Rodrigo Rafael Villarreal Goulart

Novo Hamburgo  
2019

## RESUMO

A Organização Mundial da Saúde afirma que 1 em cada 11 pessoas do mundo tem diabetes, mas somente 50% dos pacientes conhecem seu diagnóstico. O exame laboratorial mais solicitado pelos profissionais é o hemograma, que não apresenta o nível glicêmico do paciente, indicador da diabetes. Portanto é levantada a hipótese de identificar pacientes com potencial de diabetes a partir de dados de exames correlatos. Esta situação apresenta um grande desafio: detectar ou classificar o índice glicêmico do paciente através de dados hematológicos.

Nesse contexto, sistemas de apoio à decisão clínica têm demonstrado alto grau de assertividade no auxílio ao diagnóstico. A tecnologia com maior relevância da atualidade é a Inteligência Artificial, com foco principal no Aprendizado de Máquina, implementando conceitos computacionais que possibilitam aprendizado automatizado por meio de dados pré-existentes. Com foco no diagnóstico da diabetes, o presente projeto tem como objetivo a coleta e pré-processamento de dados para aplicação de técnicas de aprendizado de máquina, buscando tendências e predição da classificação do status glicêmico de pacientes, por meio de resultados de exames correlatos à glicose. A obtenção dos resultados terá como base a correlação dos resultados obtidos com a informação do real status glicêmico de cada paciente, disponível no dataset utilizado no projeto.

Palavras-chave: Inteligência Artificial; Aprendizado de máquina; Diagnóstico; Classificação glicêmica.

## SUMÁRIO

|                    |    |
|--------------------|----|
| MOTIVAÇÃO.....     | 5  |
| OBJETIVOS.....     | 8  |
| METODOLOGIA.....   | 9  |
| CRONOGRAMA .....   | 10 |
| BIBLIOGRAFIA ..... | 11 |

## MOTIVAÇÃO

Os avanços tecnológicos das últimas décadas vêm trazendo contribuições importantes para a precisão do diagnóstico médico, refletindo em benefícios para os pacientes. Este acelerado processo de inovação resulta em maior segurança na tomada de decisão clínica, tanto em casos agudos ou urgentes, quanto em doenças crônicas.

Dentro desta nova realidade, torna-se facilitado o acesso à diversos procedimentos de análise com foco no apoio ao diagnóstico, como os exames laboratoriais. A Medicina Laboratorial possui papel fundamental na tomada de decisão por parte dos médicos pois traz indicadores confiáveis sobre o estado de saúde do paciente. Exames laboratoriais são ferramentas eficazes para reduzir as incertezas da prática clínica, contribuir para a preservação e restauração da saúde e melhorar a qualidade dos cuidados de saúde (ANDRIOLO, 2008).

Diversas patologias comuns na população deixam de ser diagnosticadas ou são diagnosticadas de forma tardia devido à falta de requisições adequadas de exames clínicos pelos médicos. É sabido que diagnósticos tardios de doenças aumentam o risco de disseminação e levam a complicações, tornando o tratamento mais difícil (SHCOLNIK, 2018).

A Organização Mundial da Saúde acredita que 1 em cada 11 pessoas no mundo tem diabetes. Segundo o International Diabetes Federation (IDF), o Brasil é o quarto país com mais diabéticos no mundo, chegando a 7% da população, e o desafio passa pela falta de controle glicêmico dos pacientes: 50% dos diabéticos desconhecem o diagnóstico (SBAC, 2018).

A diabetes é uma doença metabólica: seu portador não consegue degradar moléculas de glicose corretamente ou em velocidade suficiente, sendo classificada em duas categorias: Tipo 1 e Tipo 2. A primeira é uma forma de diabetes relacionada ao sistema autoimune, em geral identificada na infância ou adolescência. As células responsáveis pela defesa do organismo acabam atacando outras, capazes de sintetizar insulina, por causa de um defeito no sistema imunológico. Na diabetes do tipo 2, o organismo não produz insulina suficiente para controlar a taxa de açúcar no sangue. A maior incidência de diabetes se concentra nesse grupo (SBAC, 2018).

Na área do diagnóstico através de exames laboratoriais, o parâmetro mais solicitado pelos profissionais da medicina é o Hemograma, seja para pacientes ambulatoriais ou hospitalizados. Segundo KAWAMOTO et al. (2013), Parâmetros hematológicos, incluindo contagem de glóbulos vermelhos, hematócrito e hemoglobina estão independentemente associados à resistência à insulina, o que caracteriza a diabetes tipo I. Em estudo publicado por KUTLU et al. (2009), afirma-se que existe uma relação significativa entre os níveis de hemoglobina e

funções celulares em pacientes com diabetes tipo 2. Diante das relações explicitadas nas pesquisas citadas, oportuniza-se a classificação do status glicêmico dos pacientes através das correlações entre variáveis hematológicas existentes.

As práticas do setor de saúde são baseadas em informações adquiridas de forma sistemática e constante, necessitando métodos de apoio na organização e formação do conhecimento com base nessas informações. O desenvolvimento das áreas de informática e de medicina abrem caminhos para estudos mais profundos sobre a relação entre elas.

A tecnologia de finalidade geral com maior relevância nessa era é a Inteligência Artificial (IA), com foco principal no Aprendizado de Máquina (AM) (BRYNJOLFSSON, 2017). O AM constitui-se de programação computacional de análise de dados para a construção de modelos analíticos. Esta programação baseia-se no conceito de que o sistema computacional pode aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana. A “máquina” pode ser protagonista no seu aprendizado e capaz de se adaptar sozinha quando exposta a novos dados (OSAKI, 2018).

Sistemas computadorizados de apoio à decisão clínica têm indicado um alto grau de acurácia em suas propostas diagnósticas. Usando diferentes algoritmos, estratégias de tomada de decisão e um grande volume de dados, sistemas de Inteligência Artificial são capazes de propor ações, quando solicitados. O Aprendizado de Máquina pode oferecer melhores indicações de riscos e de implicações da correlação entre diagnósticos e terapias, dados que poderão ser posteriormente confirmadas por estudos randomizados e controlados em uma amostra de pacientes (LOBO, 2018).

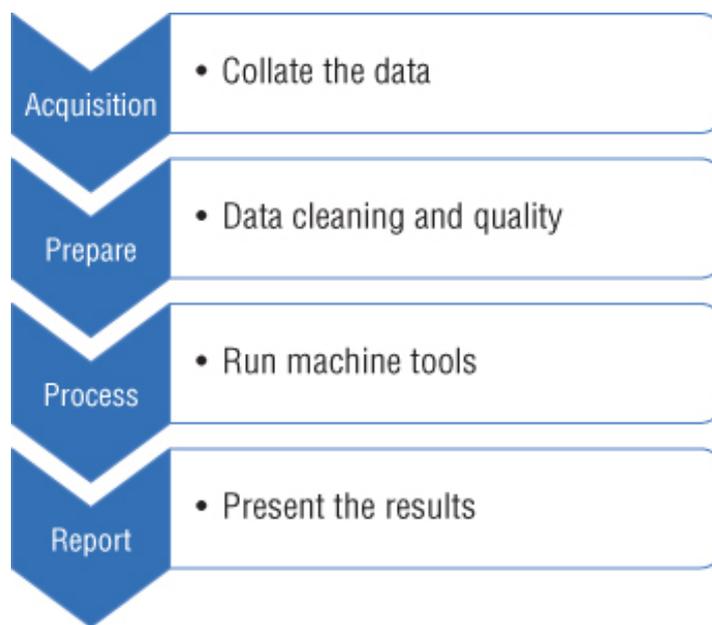
É notável o aumento e a popularização do uso de práticas de Inteligência Artificial em diversas áreas. De acordo com o relatório Tendências da Tecnologia, divulgado pela Organização Mundial de Propriedade Intelectual (OMPI) em 31 de janeiro de 2019, 50% de todas as patentes para Inteligência Artificial foram publicadas desde 2013, somando mais de 170 mil ideias (ONU – Organização das Nações Unidas, 2019).

Doutores e outros profissionais médicos usam IA para maior assertividade e agilidade nos diagnósticos. Na medicina, IA utiliza decência de dados sob informações do paciente, gerando resultados melhores que especialistas podem realizar (MURALI; SIVAKUMARAN, 2018).

O processo de obtenção de conhecimento através do AM neste estudo será composto por várias etapas, com destaque para quatro de maior relevância: 1) Coleta, análise e compreensão do conjunto de dados; 2) Aplicação das regras de pré-processamento na base de dados;

3) Aplicação das técnicas de Aprendizado de Máquina (processamento); 4) Avaliação dos resultados. Na figura 1 apresenta-se o fluxo básico de aplicação de AM.

Figura 1 – Etapas da aplicação de Aprendizado de Máquina



Fonte: BELL, 2014

Para viabilizar este estudo, o autor firmou parceria com um laboratório de análises clínicas, presente em nove cidades do estado do Rio Grande do Sul, que disponibilizará dados anônimos de diagnósticos de pacientes para composição da base de dados.

Considerando os estudos citados, diante de uma doença de abrangência mundial, com diagnóstico desconhecido ou tardio para um grande volume de pacientes, a presente pesquisa visa a coleta e pré-processamento de uma base de dados para a aplicação de técnicas de Aprendizado de Máquina, buscando tendências e predição da classificação do status glicêmico de pacientes, através de resultados de exames correlatos à glicose. A obtenção dos resultados terá como base a validação dos conhecimentos obtidos no aprendizado de máquina através da correlação com o real status glicêmico do paciente, encontrado no banco de resultados de exames de análises clínicas utilizado para a realização deste estudo.

## OBJETIVOS

### Objetivo geral

Essa pesquisa traz como objetivo geral a coleta e análise de dados de resultados de exames laboratoriais, para a aplicação de técnicas de pré-processamento e desenvolvimento de um modelo de classificador de status glicêmico dos pacientes, e sua correlação com o diagnóstico de diabetes.

### Objetivos específicos

- Extração e compreensão de banco de resultados de exames;
- Pesquisa sobre métodos de pré-processamento adequados à base de dados extraída;
- Pesquisa sobre métodos de aprendizado de máquina adequados para aplicação nos dados;
- Pré-processamento das informações do banco de resultados;
- Processamento do banco de resultados de exames e desenvolvimento de um modelo de classificador;
- Análise dos resultados obtidos e confiabilidade dos métodos aplicados;
- Avaliação dos resultados com base no real status glicêmico do paciente;

## METODOLOGIA

O estudo tem sua natureza classificada como pesquisa aplicada, pois gerará conhecimento a ser aplicado em um problema definido, originando um processo ao final. Quanto aos objetivos, tem caráter exploratório, buscando evidências sobre a questão de pesquisa durante o estudo (PRODANOV; FREITAS, 2013).

Visando a identificação dos parâmetros hematológicos correlacionados ao status glicêmico dos pacientes, será desenvolvida uma revisão bibliográfica para composição das variáveis do conjunto de dados a ser trabalhado.

Após a etapa de revisão bibliográfica será feita a análise dos dados disponíveis. A fonte de informações cedida para este estudo possui grande riqueza em detalhes de resultados quantitativos e qualitativos de exames de análises clínicas. Definida a composição do conjunto de dados, será realizada a extração destas informações diretamente do banco de dados da empresa parceira neste estudo. Os dados serão anônimos e os cuidados éticos com qualquer informação sensível serão resguardados em todas as etapas da pesquisa.

Segundo BELL (2014), os sistemas podem aprender e melhorar com a experiência e com o tempo, refinar um modelo que pode ser usado para prever resultados de perguntas com base no aprendizado anterior. Para atender aos objetivos, esta pesquisa realizará a aplicação de técnicas de pré-processamento na base de dados, a fim de adequá-la para o processamento através de Aprendizado de Máquina. Como resultado da etapa de processamento, espera-se a classificação do status glicêmico dos pacientes, com base nas variáveis hematológicas obtidas na coleta de dados.

Para validar os resultados, será incluso no conjunto de dados inicial o real status glicêmico dos pacientes. A análise dos resultados será feita por meio do confronto dos dados de status glicêmico real, com a classificação dos dados gerada pelo Aprendizado de Máquina na etapa de processamento.

Finalmente, a questão que conduz este estudo trata se é possível a classificação e previsão do status glicêmico do paciente para um pré-diagnóstico de diabetes através de dados clínicos obtidos em exames laboratoriais, por meio do Aprendizado de Máquina.

## CRONOGRAMA

## Trabalho de Conclusão I

| Etapa  | Meses |     |     |     |     |
|--|-------|-----|-----|-----|-----|
|  | Ago   | Set | Out | Nov | Dez |
| Entrega do aceite de orientação.                               |       |     |     |     |     |
| Escrita do anteprojeto.  |       |     |     |     |     |
| Revisão do anteprojeto.  |       |     |     |     |     |
| Entrega do anteprojeto.  |       |     |     |     |     |
| Definição das variáveis clínicas de estudo.                    |       |     |     |     |     |
| Extração da base de dados.                                     |       |     |     |     |     |
| Avaliação das técnicas de Machine Learning a serem utilizadas. |       |     |     |     |     |
| Redação do TCC I.  |       |     |     |     |     |
| Revisão do TCC I.  |       |     |     |     |     |
| Entrega do TCC I.  |       |     |     |     |     |

## Trabalho de Conclusão II

| Etapa  | Meses |     |     |     |     |
|--|-------|-----|-----|-----|-----|
|  | Fev   | Mar | Abr | Mai | Jun |
| Análise e aplicação das técnicas de pré-processamento. |       |     |     |     |     |
| Aplicação das técnicas de Machine Learning.            |       |     |     |     |     |
| Análise dos resultados obtidos.                        |       |     |     |     |     |
| Redação do TCC II.                                     |       |     |     |     |     |
| Revisão do TCC II.                                     |       |     |     |     |     |
| Entrega do TCC II.                                     |       |     |     |     |     |

## BIBLIOGRAFIA

- REZENDE, Joffre M. de. **O Uso da tecnologia no diagnóstico médico e suas consequências.** In: *Ética Revista – Conselho Regional de Medicina do Distrito Federal*, v. 4, n. 4, p. 18-21, 2006.
- ANDRIOLO, Adgamar. **Princípios básicos de medicina laboratorial.** In: Schor N, editor. *Guias de medicina ambulatorial e hospitalar da UNIFESP.* 2 ed. São Paulo: Manole, 2008. cap. 1, p. 1-10.
- LOBO, Luiz C. **Inteligência Artificial, o Futuro da Medicina e a Educação Médica.** In: *Revista brasileira de Educação Médica*, p. 3-8, 2018.
- OSAKI, Milton M. **Inteligência artificial, prática médica e a relação médico-paciente.** In: *Revista de Administração em Saúde*, v. 18, n. 72, 2018.
- BRYNJOLFSSON, Erik; ROCK, Daniel; Syverson, Chad. **Artificial intelligence and the modern productivity paradox: A clash of expectations and statistics.** NBER Working Paper Series, p. 24001, 2017.
- MURALI, Nivetha; SIVAKIMARAN, Nivethika. **Artificial intelligence in Healthcare - A Review.** *International Journal of Modern Computation, Information and Communication Technology*, a. 1, e. 6, p. 103-110, 2018.
- KAWAMOTO, Ryuichi et al. **Hematological parameters are associated with metabolic syndrome in Japanese community.** *International Journal of Basic and Clinical Endocrinology*, v. 43, i. 2, p. 334-341, 2013.
- BELL, Jason. **Machine Learning: Hands-on for developers and technical professionals.** Hoboken, NJ: Wiley, 2014.
- PRODANOV, Cleber Cristiano; FREITAS, Ernani Cesar de. **Metodologia do trabalho científico: métodos e técnicas da pesquisa e do trabalho acadêmico.** 2. ed. Novo Hamburgo, RS: Feevale, 2013.
- SCHOLNIK, Wilson. **Exames laboratoriais: necessidade ou desperdício?. 2019. Disponível em:** <http://www.sbpc.org.br/wp-content/uploads/2018/06/ExamesLaboratoriaisNecessidadeOuDesperdicio20mai2018.pdf>. Acesso em: 21 ago. 2019.
- SOCIEDADE BRASILEIRA DE ANÁLISES CLÍNICAS. **Qual a situação da diabetes no Brasil?. 2018. Disponível em:** <http://www.sbac.org.br/blog/2018/11/26/qual-a-situacao-da-diabetes-no-brasil/>. Acesso em: 23 ago. 2019.