

Universidade Federal de Alagoas  
Instituto de Computação



# **Predição do Absenteísmo em Agentes de Segurança Pública usando Aprendizagem Profunda**

**Edival Lima da Silva Júnior**

**Maceió-AL, Junho de 2019**

Edival Lima da Silva Júnior

# **Predição do Absenteísmo em Agentes de Segurança Pública usando Aprendizagem Profunda**

**Dissertação de Mestrado** apresentada ao Programa de Pós-Graduação em Informática da UFAL (área de concentração: Aprendizagem Profunda e Aplicações em Visão Computacional), como parte dos requisitos necessários para a obtenção do Título de Mestre em Informática.

Universidade Federal de Alagoas – UFAL

Instituto de Computação

Programa de Pós-Graduação em Informática

Orientador: Prof. Dr. Thales Miranda de Almeida Vieira

Coorientador: Prof. Dr. Evandro de Barros Costa

Maceió-AL

Junho de 2019

**Catálogo na fonte**  
**Universidade Federal de Alagoas**  
**Biblioteca Central**

Bibliotecário: Marcelino de Carvalho

S586p Silva Júnior, Edival Lima da.  
Predição do absenteísmo em agentes de segurança pública usando  
aprendizagem profunda / Edival Lima da Silva Júnior. – 2019.  
62 f. : il.

Orientador: Thales Miranda de Almeida Vieira.  
Co-orientador: Evandro de Barros Costa.  
Dissertação (mestrado em Informática) - Universidade Federal de Alagoas.  
Instituto de Computação. Maceió, 2019.

Bibliografia: f. 58-60.  
Apêndice: f. 62.

1. Absenteísmo (Trabalho). 2. Aprendizagem profunda. 3. Aprendizado do  
computador. 4. Teoria da predição. 5. Redes neurais (Computação). Título.

CDU: 004.85



UNIVERSIDADE FEDERAL DE ALAGOAS/UFAL  
Programa de Pós-Graduação em Informática – PpgI  
Instituto de Computação  
Campus A. C. Simões BR 104-Norte Km 14 BL 12 Tabuleiro do Martins  
Maceió/AL - Brasil CEP: 57.072-970 | Telefone: (082) 3214-1401



### Folha de Aprovação

Edival Lima da Silva Júnior

“Predição de Absenteísmo em Agentes de Segurança Pública usando Aprendizagem Profunda”

Dissertação submetida ao corpo docente do Programa de Pós-Graduação em Informática da Universidade Federal de Alagoas e aprovada em 19 de junho de 2019.

#### Banca Examinadora:

**Prof. Dr. Thales Miranda de Almeida Vieira**  
UFAL – Instituto de Computação  
Orientador

**Prof. Dr. Evandro de Barros Costa**  
UFAL – Instituto de Computação  
Coorientador

**Prof. Dr. Tiago Figueiredo Vieira**  
UFAL – Instituto de Computação  
Examinador Interno

**Prof. Dr. Nivan Roberto Ferreira Júnior**  
UFPE – Centro de Informática

*A Deus que nos permitiu chegar aqui  
por ser meu refúgio e fortaleza,  
apoio em todos os momentos.*

# Agradecimentos

Agradeço a Deus, por nos guiar em todas as veredas.

Aos meus pais, Edival e Aurenice e minha tia Geruza, por todo o amor, carinho e exemplo de vida simples e honesta.

À minha esposa Larissa, pelo amor, compreensão e sabedoria que me ajudou a superar todas as dificuldades.

Ao meu orientador, Prof. Dr. Thales Vieira, por todos os conselhos, dedicação, paciência e disponibilidade, além da grande ajuda nesse período.

Ao meu coorientador, Prof. Dr. Evandro Barros, pela atenção nos primeiros passos do curso, mostrando o caminho a seguir e acompanhando todo o trajeto.

Às minhas irmãs, Leilia, Lilia e Leidiane, pela paciência nas ausências e atenção e cuidado com que me confortaram.

Aos meus sogros Gilson e Berenice e minhas cunhadas Carolina e Vanessa por todo o suporte e carinho.

Aos meus amigos que souberam entender os afastamentos intermináveis e mesmo assim, permaneceram leais.

Aos todos os professores, sem eles, nada disso seria possível.

Aos novos amigos feitos no curso, Laise e Miguel e a todos os colegas que dividiram comigo essa empreitada.

*"Viver  
E não ter a vergonha  
De ser feliz  
Cantar e cantar e cantar  
A beleza de ser  
Um eterno aprendiz."  
(Gonzaguinha)*

# Resumo

O absenteísmo é um fenômeno complexo que se expressa pela ausência física do indivíduo, geralmente ao seu posto de trabalho. Nas instituições de segurança pública, essas ausências trazem inúmeros prejuízos de ordem pessoal, social e econômica, além de acontecerem em percentuais superiores aos das demais categorias profissionais. Assim, determinar seus fatores preponderantes e permitir que ações preventivas sejam realizadas de forma efetiva traria inúmeros benefícios a essas instituições e aos seus agentes. Neste trabalho, propomos e avaliamos preditores capazes de identificar os agentes mais propensos ao absenteísmo de longa duração. Esses preditores devem tomar decisões baseando-se no histórico profissional de cada agente extraído de bases de dados das instituições de segurança pública. Realizamos experimentos usando uma base de dados referente a 6 anos de afastamentos de agentes da Polícia Militar de Alagoas do Brasil, dos quais foram selecionados atributos que estariam correlacionados ao fenômeno do absenteísmo. Avaliamos arquiteturas de aprendizagem profunda do tipo *Multilayer Perceptron*, *Long Short-term Memory* e *Recurrent Neural Network*. Aplicamos técnicas de seleção de atributos e fizemos a comparação dos resultados obtidos com a técnica de Máquina de Vetores de Suporte. Apresentamos as melhores arquiteturas para predição do absenteísmo prolongado de agentes, o que comprova que é possível realizar a predição desse tipo de afastamento, atingindo 78% de acurácia, o que subsidiaria a implantação de medidas efetivas de prevenção nessas instituições. Finalmente, concluímos que o uso de dados referentes a uma quantidade maior de anos resulta em melhores resultados na predição do absenteísmo.

**Palavras-chaves:** Absenteísmo. Aprendizagem Profunda. Aprendizagem de Máquina. Predição.

# Abstract

Absenteeism is a complex phenomenon that is expressed by the physical absence of the individual, usually at his job. In public security institutions, these absences bring many personal, social and economic losses, in addition to occurring in percentages superior to those of the other professional categories. Thus, determining its preponderant factors and allowing preventive actions to be carried out effectively would bring numerous benefits to these institutions and their agents. In this work, we propose and evaluate predictors capable of identifying the most prone agents to long-term absenteeism. These predictors should make decisions based on the professional history of each agent extracted from databases of public security institutions. We carried out experiments using a database comprised of 6 years of absences from agents of the Military Police of Alagoas in Brasil, from which we selected attributes that would be correlated to the phenomenon of absenteeism. We evaluated Deep Learning architectures such as Multilayer Perceptron, Long Short-term Memory and Recurrent Neural Network. We applied attributes selection techniques and compared the results obtained by the Machine Learning Support-Vector Machine technique. We present the best architectures for predicting the prolonged absenteeism of agents, which proves that it is possible to predict this type of absence, reaching 78% of accuracy, which would support the implementation of effective prevention measures in these institutions. Finally, we conclude that the use of data referring to a more significant number of years results in better results in the prediction of absenteeism.

**Keywords:** Absenteism. Deep Learning. Machine Learning. Prediction

# Lista de ilustrações

Figura 1	– Hierarquia das categorias de absentéismo adaptado do Relatório da AFOEM (1999) para destacar a hierarquia dos afastamentos motivados por adoecimento nas organizações. . . . .	8
Figura 2	– Um diagrama de <i>Venn</i> mostrando como a aprendizagem profunda é um tipo de aprendizado por representação, que é, por sua vez, um tipo de aprendizado de máquina, que é usado para muitas, mas nem todas as abordagens de Inteligência Artificial. Cada seção do diagrama de <i>Venn</i> inclui um exemplo de uma tecnologia de IA. . . . .	10
Figura 3	– Diagrama em blocos da aprendizagem com um professor . . . . .	12
Figura 4	– Célula neural biológica com a sequência de propagação do sinal . . . . .	13
Figura 5	– Modelo não-linear de um neurônio artificial recebendo as entradas $x_j$ , sendo processado pelo neurônio $k$ , com uma função de ativação $\varphi$ e tendo como saída $y_k$ . . . . .	14
Figura 6	– Adaptação dos gráficos típicos das funções de ativação relacionadas por Michael (2005) como as mais bem sucedidas em aplicações práticas: (a) retificadora; (b) função por limiar; (c) função sigmoide; (d) função tangente. . . . .	16
Figura 7	– Um exemplo hipotético de uma rede <i>Multilayer Perceptron</i> . . . . .	22
Figura 8	– Estrutura esquematizada de uma célula RNN e suas conexões recorrentes nos tempos $t - 1$ , $t$ e $t + 1$ . . . . .	23
Figura 9	– Estrutura esquematizada de uma célula LSTM e suas conexões recorrentes nos tempos $t - 1$ , $t$ e $t + 1$ . . . . .	24
Figura 10	– Exemplo de aplicação de SVM linear para classificação dos elementos de um conjunto de treinamento . . . . .	25
Figura 11	– Exemplo de aplicação de SVM com $Kernel_{RBF}$ para classificação dos elementos de um conjunto de treinamento . . . . .	26
Figura 12	– Processo de extração dos dados dos sistemas de Recursos Humanos e de Gestão Hospitalar da PMAL . . . . .	28
Figura 13	– Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 2 anos acumulados (a); para 3 anos acumulados (b) - 2012-2017 . . . . .	41
Figura 14	– Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 4 anos acumulados (a); para 5 anos acumulados (b) - 2012-2017 . . . . .	41
Figura 15	– Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 6 anos acumulados (a); para todos os intervalos de anos acumulados (b) - 2012-2017 . . . . .	42

Figura 16 – Gráfico da acurácia das melhores arquiteturas, considerando a quantidade de anos acumulados: eixo horizontal representa a quantidade de anos e o eixo vertical a melhor acurácia - 2012-2017 . . . . .	43
Figura 17 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 2 anos em sequência - 2012-2017 . . . . .	44
Figura 18 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 3 anos em sequência - 2012-2017 . . . . .	44
Figura 19 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 4 anos em sequência - 2012-2017 . . . . .	44
Figura 20 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 5 anos em sequência - 2012-2017 . . . . .	45
Figura 21 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 6 anos em sequência - 2012-2017 . . . . .	45
Figura 22 – Projeção t-SNE 2D para absentes e não absentes para dados de 6 anos em sequência - 2012-2017 . . . . .	47
Figura 23 – Projeção t-SNE 3D para absentes e não absentes para 6 anos em sequência - 2012-2017 . . . . .	48
Figura 24 – Projeção t-SNE 2D para absentes e não absentes para 6 anos em sequência com redução de dimensionalidade por PCA - 2012-2017 . . . . .	49
Figura 25 – Frequência percentual relativa de absenteísmo de longa duração considerando o número de afastamentos superiores a 30 dias de toda a instituição - 2012-2017	52
Figura 26 – Frequência percentual relativa de absenteísmo de longa duração considerando o número de afastamentos superiores a 30 dias em relação ao efetivo de cada unidade - 2012-2017 . . . . .	52
Figura 27 – Comparação entre curva ROC e AUC das melhores configurações de arquiteturas e seleção de atributos pelo método <i>Wrapper</i> para 6 anos em sequência - 2012-2017 . . . . .	54

# Lista de tabelas

Tabela 1 – Exemplo de dados de agentes agregados para 2 anos em sequência - 2012-2017	30
Tabela 2 – Frequência de conjuntos de treinamento por ano em sequência rotulados como ausentes ou não ausentes - 2012-2017 . . . . .	30
Tabela 3 – Quantitativo de experimentos por arquitetura e de variações de hiperparâmetros - 2018 . . . . .	31
Tabela 4 – Melhores arquiteturas para as redes do tipo MLP por anos em sequência - 2012-2017 . . . . .	35
Tabela 5 – Melhores índices de acurácia para arquiteturas do tipo MLP - 2012-2017 . .	36
Tabela 6 – Configurações das melhores arquiteturas para as redes do tipo RNN por anos em sequência - 2012-2017 . . . . .	36
Tabela 7 – Melhores índices de acurácia para arquiteturas do tipo RNN - 2012-2017 . .	37
Tabela 8 – Configurações das melhores arquiteturas para as redes do tipo LSTM por anos em sequência - 2012-2017 . . . . .	38
Tabela 9 – Melhores índices de acurácia para arquiteturas do tipo LSTM - 2012-2017 .	38
Tabela 10 – Configurações das melhores arquiteturas para as SVM por anos em sequência - 2012-2017 . . . . .	39
Tabela 11 – Melhores índices de acurácia para arquiteturas do tipo SVM - 2012-2017 . .	39
Tabela 12 – Comparação da acurácia entre as melhores arquiteturas de MLP, LSTM, RNN e SVM por anos em sequência - 2012-2017 . . . . .	40
Tabela 13 – Resultado de acurácia para a seleção de atributos por um método <i>Wrapper</i> para MLP de 5 camadas e tendo como entrada os dados de 6 anos em sequência - 2012-2017 . . . . .	50
Tabela 14 – Frequência de características para o método de seleção de atributos <i>Wrapper</i> para os 10 melhores resultados retornados para MLP de 5 camadas e tendo como entrada os dados de 6 anos em sequência - 2012-2017 . . . . .	51

# Lista de abreviaturas e siglas

AC	Acurácia
AFOEM	<i>Australasian Faculty of Occupational and Environmental Medicine</i>
ANO	Ano do registro do afastamento
ANO_SERV	Anos de efetivo serviço
AUC	Área sob uma curva de característica de operação de um receptor
BPE	Batalhão de Policiamento de Eventos
BPGD	Batalhão de Polícia de Guardas
BPM	Batalhão de Polícia Militar
BPRP	Batalhão de Policiamento de Rádio Patrulha
Cap	Capitão
Cb	Cabo
CCSV	Companhia de Comando e Serviço
COND	Condição do agente no serviço ativo
CORR	Corregedoria
CPM	Companhia de Polícia Militar
CSV	<i>Comma Separated Values</i>
DE	Diretoria de Ensino
DP	Diretoria de Pessoal
EMG	Estado Maior Geral
FN	Falso Negativo
FP	Falso Positivo
GÊNERO	Gênero do agente
IA	Inteligência Artificial

ID	Identificador do Policial Militar no sistema de Recursos Humanos
IDADE	Idade no momento do afastamento
LSTM	<i>Long Short-term Memory</i>
MAE	<i>Mean Absolute Error</i>
Maj	Major
MLP	<i>Multilayer Perceptron</i>
NÍVEL	Nível hierárquico dentro da instituição
OPM	Organização Policial Militar
OPM	Unidade onde o agente exerce suas atividades
PCA	<i>Principal Component Analysis</i>
PM4	4ª Seção do Estado Maior Geral
PMAL	Polícia Militar de Alagoas
QTD	Quantidade
RBF	<i>Radial Basis Function</i>
RELU	<i>Rectified Linear Unit</i>
RNN	<i>Recurrent Neural Network</i>
ROC	Curva característica de operação de um receptor
SEDS	Secretaria de Estado de Defesa Social (atual Secretaria de Segurança Pública)
SEQ	Sequência
SGBD	Sistema de Gerenciamento de Banco de Dados
Sgt	Sargento
SisHosp	Sistema de Gestão Hospitalar
SOM_ABS	Somatório de dias de afastamento por adoecimento
SVM	Máquina de Vetores de Suporte
t-SNE	<i>T-Distributed Stochastic Neighbor Embedding</i>
TVN	Taxa de Verdadeiro Negativo

TVP	Taxa de Verdadeiro Positivo
VN	Verdadeiro Negativo
VP	Verdadeiro Positivo

# Lista de símbolos

$\gamma$	Letra grega minúscula Gama
$\Phi$	Letra grega maiúscula Fi
$\varphi$	Letra grega minúscula Fi
$\Sigma$	Letra grega maiúscula Sigma
$\sigma$	Letra grega minúscula Sigma
$\Theta$	Letra grega maiúscula Teta
$\varepsilon(t)$	Função de custo ou índice de desempenho no tempo de iteração $t$
$\varepsilon$	Letra grega minúscula Epsilon
$A$	Célula da arquitetura LSTM ou RNN
$b_k$	Bias $b$ atribuído a um neurônio $k$
$C_t$	Estado oculto da célula LSTM ou RNN no tempo $t$
$C$	Hiperparâmetro de regularização da margem da função de decisão de uma SVM
$e_k(t)$	Sinal de erro $e$ no tempo de iteração $t$ para um neurônio $k$
$f(x)$	Função de ativação $f$ para uma entrada $x$
$g(z)$	Função de ativação $g$ para uma entrada $z$
$h_{\langle t \rangle}$	Saída da arquitetura no tempo $t$
$h_{\langle t-1 \rangle}$	Saída da arquitetura no tempo $t - 1$
$h_{\langle t+1 \rangle}$	Saída da arquitetura no tempo $t + 1$
$h$	Saída de uma célula LSTM ou RNN
$K$	Kernel de uma SVM não linear
$\max\{z\}$	Máximo valor retornado por uma função de ativação para uma entrada $z$
$t - 1$	Passo de iteração no tempo $t - 1$
$t + 1$	Passo de iteração no tempo $t + 1$

$t$	Passo de iteração no tempo $t$
$\tanh$	Função de ativação tangente hiperbólica
$u_k$	Campo local induzido de uma entrada $u$ para um neurônio $k$
$w_{kj}$	Peso sináptico $w$ de um neurônio $k$ por meio de um terminal de entrada $j$
$x_{t-1}$	Entrada da arquitetura no tempo $t - 1$
$x_t$	Entrada da arquitetura no tempo $t$
$x_{t+1}$	Entrada da arquitetura no tempo $t + 1$
$x(t)$	Vetor de entrada da arquitetura $x$ no tempo de iteração $t$
$y_k(t)$	Saída da arquitetura $y$ de um neurônio $k$ no tempo de iteração $t$
$y_k$	Saída $y$ de um neurônio $k$
$y'$	Saída discretizada de uma arquitetura
$y$	Saída de uma arquitetura

# Sumário

<b>1</b>	<b>INTRODUÇÃO</b>	<b>1</b>
<b>1.1</b>	<b>Motivação</b>	<b>3</b>
<b>1.2</b>	<b>Objetivos</b>	<b>5</b>
1.2.1	Objetivo Geral	5
1.2.2	Objetivos Específicos	5
<b>1.3</b>	<b>Trabalhos Relacionados</b>	<b>5</b>
<b>1.4</b>	<b>Estrutura do Trabalho</b>	<b>6</b>
<b>2</b>	<b>FUNDAMENTAÇÃO TEÓRICA</b>	<b>8</b>
<b>2.1</b>	<b>Absenteísmo</b>	<b>8</b>
<b>2.2</b>	<b>Aprendizagem Profunda</b>	<b>9</b>
<b>2.3</b>	<b>Aprendizado Supervisionado</b>	<b>11</b>
<b>2.4</b>	<b>Redes Neurais Artificiais</b>	<b>13</b>
2.4.1	Estrutura de um neurônio artificial	14
2.4.2	Funções de ativação	15
2.4.2.1	Função de ativação do tipo retificadora	15
2.4.2.2	Função de ativação por limiar	16
2.4.2.3	Função de ativação do tipo sigmoide	16
2.4.2.4	Função de ativação do tipo tangente hiperbólica	17
2.4.3	Função de custo	17
2.4.4	Algoritmo <i>Back-propagation</i>	18
2.4.5	Algoritmo de otimização	19
<b>2.5</b>	<b>Arquiteturas utilizadas</b>	<b>20</b>
2.5.1	Arquitetura do tipo <i>Multilayer Perceptron</i>	21
2.5.2	Arquitetura do tipo <i>Recurrent Neural Network</i>	22
2.5.3	Arquitetura do tipo <i>Long Short-term Memory</i>	23
2.5.4	Máquina de Vetores de Suporte	24
<b>3</b>	<b>MATERIAIS E MÉTODOS</b>	<b>27</b>
<b>3.1</b>	<b>Pré-processamento dos dados</b>	<b>27</b>
3.1.1	Agregação dos conjuntos de treinamento	29
3.1.2	Configuração da variabilidade de hiperparâmetros para as arquiteturas selecionadas	31
3.1.2.1	<i>Multilayer Perceptron</i>	33
3.1.2.2	<i>Recurrent Neural Network</i>	33
3.1.2.3	<i>Long Short-term Memory</i>	33

3.1.2.4	Máquina de Vetores de Suporte . . . . .	34
<b>3.2</b>	<b>Detalhes de implementação</b> . . . . .	<b>34</b>
<b>4</b>	<b>RESULTADOS E DISCUSSÃO</b> . . . . .	<b>35</b>
<b>4.1</b>	<b>Resultados</b> . . . . .	<b>35</b>
4.1.1	<i>Multilayer Perceptron</i> . . . . .	35
4.1.2	<i>Recurrent Neural Network</i> . . . . .	36
4.1.3	<i>Long Short-term Memory</i> . . . . .	37
4.1.4	Máquina de Vetores de Suporte . . . . .	39
4.1.5	Comparação entre os resultados obtidos pelas arquiteturas selecionadas	40
4.1.6	Visualização dos dados pela técnica de projeção multidimensional <i>T-Distributed Stochastic Neighbor Embedding</i> . . . . .	46
4.1.7	Análise da influência e relação entre atributos por um método <i>Wrapper</i>	49
4.1.8	Análise comparativa das arquiteturas por Curvas Características de Operação dos Receptores . . . . .	53
<b>4.2</b>	<b>Discussão</b> . . . . .	<b>55</b>
	<b>Conclusão e Trabalhos Futuros</b> . . . . .	<b>56</b>
	<b>REFERÊNCIAS</b> . . . . .	<b>58</b>
	<b>APÊNDICES</b>	<b>61</b>
	<b>APÊNDICE A – REPOSITÓRIO GITHUB</b> . . . . .	<b>62</b>

# 1 Introdução

O absenteísmo é um fenômeno que traz inúmeros prejuízos às organizações, inclusive àquelas responsáveis pela segurança pública, afligindo seus agentes de forma contumaz. Desse modo, o tema é alvo de pesquisas oriundas de diversas áreas, principalmente aquelas voltadas à psicologia, economia, medicina, administração e recentemente, a informática. Essas ciências alicerçaram alguns conceitos sobre o tema, como destacam [Gosselin, Lemyre e Corneil \(2013\)](#) quando citam diversos autores que estabeleceram parâmetros para definir o fenômeno. Desses conceitos, um dos mais relevantes define o absenteísmo como “o comportamento que leva à falta de presença física em um ambiente onde se esperava que o indivíduo estivesse”, geralmente associado ao ambiente de trabalho.

Nesse mesmo estudo, os autores destacam a existência de diversas teorias para explicar as variáveis que estariam envolvidas com o absenteísmo no trabalho, adotando como referência o paradigma da escolha da decisão em que o indivíduo consciente ou inconscientemente opta por fazer-se ou não presente no ambiente laboral. Dentre esses fatores, cinco são alvo de maiores estudos por parte dos pesquisadores, como bem destacam os autores, são eles: indicadores sociodemográficos, personalidade, comportamento no local de trabalho, contexto social e o processo de decisão em si, como também estariam associados a problemas de saúde específicos ([GOSSELIN; LEMYRE; CORNEIL, 2013](#)).

Conseqüentemente, as ausências de agentes de segurança pública nas instituições causadas pelo absenteísmo trazem inúmeros prejuízos de ordem social e econômica pela diminuição do número de profissionais que deveriam exercer seu mister de garantia da segurança pública. Assim, além desse prejuízo evidente, há a perda da capacidade laboral que se agrava quando as causas desse fenômeno não são diagnosticadas precocemente, levando a afastamentos por períodos cada vez mais longos.

Outro fenômeno relacionado ao absenteísmo e que opera de forma diversa é o presenteísmo, que no trabalho de [Nicolas et al. \(2018\)](#) caracteriza-se pela presença física no trabalho apesar de possuir um problema de saúde que justificaria sua ausência. Esse tipo de comportamento, como sugere o estudo dos autores, traz também danos aos profissionais que o praticam, pois revelou que 40% dos trabalhadores com sintomas depressivos relacionados ao trabalho fazem presenteísmo de longa duração, ou seja, permanecem indo ao trabalho quando deveriam restabelecer primeiramente a saúde. Assim, os agentes de segurança que não são submetidos a tratamentos preventivos, podem continuar exercendo suas atividades, quando deveriam estar submetidos a tratamentos adequados, o que se não for colocado em prática, em casos mais graves, podem vir a se afastar por longos períodos ou mesmo não retornar a exercer suas atividades, sendo afastados definitivamente.

Para [Nicolas et al. \(2018\)](#), destacam-se diversos fatores que contribuiriam para o presenteísmo e por consequência com o absenteísmo posterior, quais sejam, as políticas internas de gestão de assiduidade, as condições de trabalho, a insegurança no emprego, as dificuldades de substituição, sentimentos de injustiça no trabalho, a carga de trabalho, a autonomia de decisão, os tipos de problemas psíquicos dos indivíduos, a duração de um problema físico, as condições físicas de uma estação de trabalho, dentre outros, sendo essas características facilmente transportadas para a realidade dos agentes de segurança pública.

Desse modo, essa realidade se tornou ainda mais evidente quando da realização da pesquisa de saúde, segurança no trabalho e qualidade de vida dos operadores de segurança pública de Alagoas, realizada pela Universidade Federal de Alagoas sob patrocínio da Secretaria de Segurança Pública do Estado de Alagoas. Nela foram apontados diversos fatores que poderiam contribuir para a degradação da saúde dos agentes de segurança pública e que seriam motivadores do adoecimento e afastamento cada vez mais frequentes desses profissionais ([SILVA, 2015](#)).

Assim, sabendo-se que o absenteísmo é um fenômeno composto por um complexo de elementos relacionados e que nas palavras de Johns, citado por [Gosselin, Lemyre e Corneil \(2013\)](#), requer uma substancial investigação, optamos por estudar esse fenômeno através dos dados relacionados aos afastamentos disponíveis nas instituições, que permitiriam traçar um paralelo entre as características comuns aos indivíduos que se afastam e as dos que não se afastam do ambiente de trabalho, identificando, inclusive, aqueles que estariam praticando o presenteísmo, mesmo reunindo condições de se afastar de suas atividades para serem submetidos a tratamentos necessários. Esse tipo de comportamento, o presenteísmo, geralmente antecede o absenteísmo e pode dificultar a identificação dos profissionais que se afastarão por longos períodos ([NICOLAS et al., 2018](#)). Esses dados estão disponíveis nas instituições em seus sistemas, porém, não estão sendo utilizados como um meio eficaz de se extrair esse conhecimento.

Para esse tipo de problema, de elevada complexidade, em que a informação está disponível em uma extensa massa de dados, as redes neurais artificiais têm sido utilizadas com muito sucesso na detecção de modelos que representem distribuições de probabilidades, principalmente aquelas voltadas à análise de dados que podem ser classificados em instâncias de uma determinada categoria, utilizando-se de algoritmos de aproximação ([GOODFELLOW et al., 2014](#)). No problema em análise, serão utilizadas arquiteturas de redes neurais profundas para gerar o conhecimento de quais fatores levariam ao elevado número de afastamentos de agentes de segurança pública, busca essa que seria impossível sem o auxílio dessas técnicas.

Logo, no contexto em análise, essas técnicas foram utilizadas inicialmente para determinar que características levariam ao elevado índice de absenteísmo entre os agentes de segurança pública, permitindo identificar precocemente e com alta probabilidade, os agentes propensos a se afastarem por longos períodos. É importante destacar que a construção de modelos preditivos para esse problema que apresentem acurácia moderada, já será de extrema relevância, considerando-se a imprevisibilidade do problema e a possibilidade de focar medidas preventivas

nos profissionais indicados pelo modelo.

Assim, esperamos que o presente trabalho lance luz sobre a aplicabilidade de técnicas de aprendizagem profunda na solução de problemas do dia a dia, principalmente naqueles em que a complexidade das variáveis envolvidas, como é a natureza do comportamento humano, possa resultar em ferramentas capazes de subsidiar atividades de prevenção que permitam a melhoria da qualidade de vida de profissionais de diversas áreas, em específico, aqueles envolvidos na manutenção da segurança pública, sem a qual não é possível desenvolver quaisquer outras atividades.

## 1.1 Motivação

As instituições de segurança pública estão presentes no cotidiano da população e dispostas em todos os recantos do território brasileiro, sendo uma das categorias profissionais mais presentes junto aos cidadãos, garantindo a paz e a ordem social como preconiza o Art. 144 da Constituição Federativa do Brasil (BRASIL, 1988). Os agentes de segurança pública representam o público interno dessas instituições que é composto por mulheres e homens que se revezam diuturnamente para a realização das mais variadas atividades relacionadas à segurança, como policiais civis, policiais militares, peritos oficiais, agentes penitenciários, bombeiros militares, policiais federais, policiais rodoviários federais, policiais ferroviários federais, guardas municipais, agentes de trânsito etc, e garantem que todas as demais categorias profissionais possam exercer seu mister.

Contudo, após a realização da pesquisa de saúde, segurança no trabalho e qualidade de vida dos operadores de segurança pública de Alagoas, constatou-se que esses profissionais estão se afastando em índices alarmantes, como deixa claro o estudo, no qual 34,4% dos agentes de segurança pública entrevistados fizeram ou fazem uso de dispensas ou licenças para tratamento de saúde. Esse índice chega aos 52,96% quando analisados os agentes que compõem o Corpo de Bombeiros Militares de Alagoas isoladamente (SILVA, 2015).

Assim, é clara a necessidade de se estudar as causas desses afastamentos, que fazem com que as instituições não atinjam seus objetivos de forma satisfatória, perdendo uma fração considerável de seu efetivo para causas que até aquele momento não haviam sido estudadas com a profundidade necessária. Além disso, esses profissionais, após reiteradas curtas ausências, são submetidos às juntas médicas de saúde apenas quando as causas já estão sedimentadas e quando os afastamentos se tornam prolongados, dificultando o tratamento e recuperação desses profissionais.

Outrossim, esses profissionais que não estão presentes no dia a dia da sociedade contribuem para que as instituições não entreguem seu principal produto, que é a sensação de segurança, que permite que qualquer outra atividade da sociedade seja posta a efeito. Com isso, fica patente que o capital humano é um dos recursos mais valiosos dessas instituições, e como

tal, deve ser valorizado e preservado a fim de garantir que toda a sociedade usufrua da paz e da tranquilidade para a realização de suas outras atividades.

Percebe-se também que essa problemática não é exclusiva das forças de manutenção da segurança pública no Brasil. Em [Corneil et al. \(1999\)](#), estudou-se a prevalência de sintomas de estresse pós-traumático nos corpos de bombeiros dos Estados Unidos e no Canadá, observou-se a incidência de elevados níveis de adoecimento, encontrando como justificativas para o fenômeno a exposição desses profissionais a riscos e situações estressantes inerentes ao desempenho de suas atividades e a falta de apoio nas instituições.

Em [Minayo, Assis e Oliveira \(2011\)](#), estudou-se a influência do tipo de atividade exercida pelos policiais civis e militares do Rio de Janeiro e as condições de saúde desses profissionais, demonstrando que o risco inerente a essa atividade tem relação direta com o elevando índice de adoecimento e conseqüentemente afastamentos de suas atividades, e que esses afastamentos acontecem em índices muito superiores aos encontradas em outros grupos populacionais. Segundo o trabalho, no último ano da pesquisa, 21,6% dos policiais militares e 6,7% dos policiais civis tiveram que ser afastados de suas atividades por incapacidade temporária associada ao exercício profissional.

Com isso, buscamos uma forma de se evitar ou minimizar esses afastamentos, agindo de forma preventiva e auxiliando os órgãos responsáveis pelo acompanhamento do estado de saúde desses profissionais, para que possuam ferramentas adequadas ao enfrentamento do problema. Assim, constatamos que existe uma massa de dados não aproveitada em diversos sistemas das instituições. Esses sistemas guardam em seus registros dados importantes de cada profissional e que poderiam ser estudadas para a criação de ferramentas tecnológicas que permitissem inferir características comuns aos profissionais que estariam se afastando por longos períodos de tempo, e com base nessas características, promover ações preventivas e direcionadas a esse público. Outrossim, essas tecnologias ainda não são utilizadas pelos órgãos de segurança pública estaduais, municipais ou mesmo federais.

Fica demonstrado que as instituições públicas de segurança perdem a cada dia uma fração importante de indivíduos que são afastados de suas atividades laborais por motivo de adoecimento, seja ele causado por fatores físicos ou psicológicos, que além de trazerem prejuízos à sociedade que tanto necessita desses profissionais, resultam em danos muitas vezes irremediáveis ao indivíduo que não consegue mais realizar as atividades para as quais foi incumbido.

Assim, a problemática do presente trabalho trata da seguinte hipótese de pesquisa: seria possível a construção de um preditor baseado em técnicas de aprendizagem profunda capaz de extrair características de profissionais de segurança pública que se afastam ou não por longos períodos, baseando-se nos dados presentes em diversos sistemas das instituições e que permitisse o direcionamento de atividades preventivas àqueles que estiverem mais propensos ao absenteísmo por adoecimento?

Logo, no presente trabalho, pretendemos demonstrar que isso é possível por meio de uma solução tecnológica, baseada em aprendizagem profunda, que permitirá aos setores responsáveis pela saúde física e psicológica dos agentes de segurança pública nas diversas instituições, agirem preventivamente a fim de ofertar um cuidado direcionado àqueles agentes que apresentam características comuns aos que se afastam por longos períodos, de modo que haja uma ação eficaz e efetiva e de certo modo economicamente viável para que o fenômeno do absenteísmo motivado pelo adoecimento seja enfrentado de forma assertiva.

## 1.2 Objetivos

### 1.2.1 Objetivo Geral

a) Demonstrar que técnicas de aprendizagem profunda são capazes de prever o absenteísmo entre agentes de segurança pública motivado por adoecimento.

### 1.2.2 Objetivos Específicos

a) Avaliar diversas classes de arquiteturas de aprendizagem profunda para verificar quais seriam mais adequadas à identificação de profissionais sujeitos a afastamentos prolongados;

b) Adequar os dados presentes em diversos sistemas informatizados das instituições para que possam ser usados como insumos no desenvolvimento de redes neurais artificiais capazes de prever o afastamento de profissionais de segurança pública;

c) Identificar possíveis relações entre características que estariam envolvidas no elevado índice de absenteísmo nas instituições;

d) Criar modelos de predição capazes de identificar com antecedência profissionais que possam vir a ser afastados, permitindo a ação preventiva dos setores responsáveis pela saúde física e psicológica nas instituições.

## 1.3 Trabalhos Relacionados

O absenteísmo é objeto de diversas pesquisas, sendo inúmeras as motivações para esses estudos, destacando-se principalmente aquelas voltadas ao entendimento das circunstâncias que levam à ausência do indivíduo em seu ambiente de trabalho, haja vista o prejuízo experimentado pelas organizações e principalmente ao indivíduo que vê sua produtividade decair quando não são identificadas as causas desse comportamento.

Em [Gosselin, Lemyre e Corneil \(2013\)](#) destaca-se que o tema é foco de estudos por mais de 40 anos, e que se acreditava que todos os indivíduos produziam de forma igual, mas o que ocorre é que as organizações experimentam a crescente incidência de faltas de longa duração

e a perda de produtividade gerada por essas ausências. Assim, propõem o estudo do tema por meio de um questionário para avaliar a incidência de diversas características nos indivíduos que se afastam por longos períodos, como características individuais, demográficas, fatores organizacionais e problemas de saúde em específico, utilizando-se de uma matriz de correlação para aferir e demonstrar seus resultados. Em nosso trabalho, também procuramos características relevantes, porém usando técnicas de Aprendizagem Profunda.

Já em [Fan et al. \(2012\)](#) o tema é estudado por meio da aplicação de técnicas híbridas de Aprendizagem de Máquina por meio de *clusters* para realizar a predição de afastamentos de profissionais da área de tecnologia. Entretanto, a pesquisa é baseada em questionários distribuídos após o ano novo chinês que seria o período de maior incidência desse tipo de afastamento em Taiwan. Conseqüentemente, o resultado depende da disposição do avaliado em responder corretamente aos questionamentos, o que pode influenciar os resultados obtidos. Contudo, demonstrou-se a possibilidade de previsão desse fenômeno pela análise de características dos indivíduos e das organizações.

Em [Homaie-Shandizi et al. \(2016\)](#), o absenteísmo foi tratado como importante fator determinante de custos em companhias aéreas, em que a ausência de um ou mais integrantes da tripulação pode acarretar prejuízos e atrasos nesse tipo de indústria. Buscou nessa pesquisa realizar a predição do absenteísmo pelo emprego de árvores de decisão sobre dados históricos de ausências da tripulação com a finalidade de permitir um correto dimensionamento da tripulação reserva, que também representa um elevado custo para as companhias aéreas. Contudo, o trabalho considera apenas fatores ligados diretamente às companhias aéreas, como o agendamento prévio da tripulação e histórico de faltas, sem relacioná-los às características intrínsecas dos integrantes da tripulação e da própria companhia, que são fatores envolvidos e motivadores do absenteísmo.

No presente estudo, buscamos realizar a predição do absenteísmo de longa duração motivado por adoecimento utilizando como subsídio os dados disponíveis nos sistemas das instituições, que serão usados para treinar arquiteturas de redes neurais que serão capazes, com base no histórico de afastamentos, de características intrínsecas dos agentes e de outras características relacionadas às instituições de segurança pública, de classificá-los como potenciais indivíduos sujeitos ao absenteísmo de longa duração iminente. Isto permitirá o conseqüente encaminhamento precoce destes profissionais aos serviços de prevenção e acompanhamento de saúde física e mental nas instituições.

## 1.4 Estrutura do Trabalho

Nosso trabalho seguiu a seguinte estrutura a fim de permitir uma rápida compreensão do problema e da solução proposta.

- **Capítulo 2 - Fundamentação Teórica** - Neste capítulo apresentamos uma breve concei-

tuação sobre o tema, uma descrição das técnicas de aprendizagem profunda utilizadas e suas principais características e fizemos um comparativo com a técnica de Máquina de Vetores de Suporte para contrapor às técnicas de Aprendizagem Profunda.

- **Capítulo 3 - Materiais e Métodos** - Neste capítulo descrevemos quais foram os materiais e métodos utilizados para o atingimento dos objetivos propostos, quais foram os hiperparâmetros investigados em cada arquitetura e como se pretende demonstrar a eficácia da metodologia adotada.
- **Capítulo 4 - Resultados e Discussão.** Neste capítulo apresentamos os resultados obtidos pelas diversas arquiteturas experimentadas. Fizemos uma comparação entre os resultados obtidos por cada arquitetura, realizamos a seleção de atributos para avaliar quais estariam mais relacionados ao fenômeno e apresentamos uma visualização dos dados por t-SNE que demonstra a complexidade do problema em estudo.
- **Capítulo 5 - Conclusão e Trabalhos Futuros** - Neste capítulo apresentamos as nossas conclusões e quais seriam as implementações possíveis para a aplicação da proposta de forma prática no dia a dia das instituições.

## 2 Fundamentação Teórica

Nesse capítulo faremos uma breve conceituação do que é o absenteísmo e das principais técnicas que utilizamos no trabalho, bem como quais são as principais características de cada técnica empregada.

### 2.1 Absenteísmo

Segundo o relatório da [AFOEM \(1999\)](#), não existe uma definição formal e universal para o absenteísmo em relação ao local de trabalho, sendo um tanto complicada a interpretação da literatura produzida pelo tema, o que dificultaria, inclusive, a mensuração e comparação entre instituições e que a adoção de padrões para o registro por governos e indústrias ajudaria muito no estudo do fenômeno.

Contudo, para [Gosselin, Lemyre e Corneil \(2013\)](#), o absenteísmo poderia ser conceituado genericamente, como a falta de presença física em um cenário onde o comportamento esperado do indivíduo seria que ele estivesse, assim, assevera que apesar do interesse acentuado no tema e de inúmeras estratégias de intervenção, o absenteísmo ainda é um problema organizacional em muitos países.

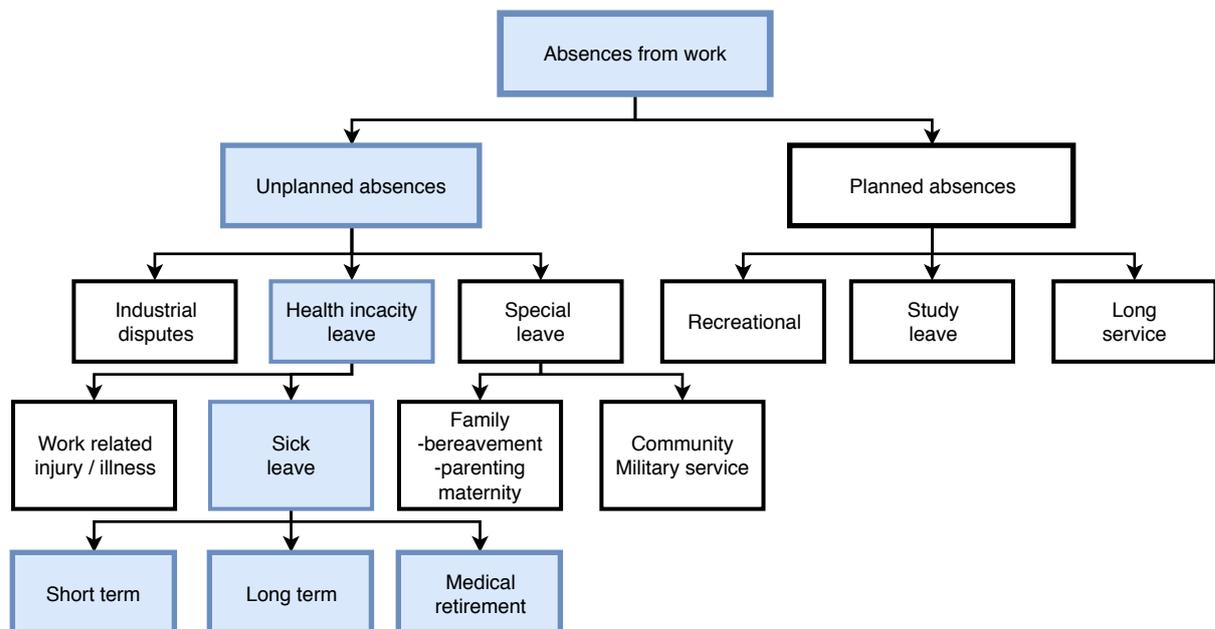


Figura 1 – Hierarquia das categorias de absenteísmo adaptado do Relatório da [AFOEM \(1999\)](#) para destacar a hierarquia dos afastamentos motivados por adoecimento nas organizações.

Fonte: [AFOEM \(1999\)](#)

Do relatório da [AFOEM \(1999\)](#), extraímos que, as ausências ao trabalho podem ser classificadas como planejadas, que seriam as férias, recreacional, licença de estudo, compensação por jornadas prolongadas etc. E as não planejadas, que seriam as licenças maternidade, paternidade, luto, problemas familiares e aquelas que ocorrem por motivo de doenças físicas ou psicológicas, dentre outras, sendo difícil formular um rol taxativo dessas motivações, mas que poderiam ser esquematizadas conforme Figura 1.

Dentro dessa hierarquia, vemos que as ausências não planejadas (*Unplanned absences*) dão origem às licenças por incapacidade de saúde (*health incapacity leave*), que são amparadas por atestados médicos (*sick leave*), que dão origem aos afastamentos de curta, longa ou mesmo às aposentadorias médicas (*medical retirement*), fazendo também distinção entre os afastamentos por doenças relacionadas ao trabalho e por lesões relacionadas ao trabalho (Figura 1).

Desse modo, preferimos a conceituação estabelecida por [Gosselin, Lemyre e Corneil \(2013\)](#) em que a falta de presença física ao ambiente laboral é suficiente para caracterizar o absenteísmo e do relatório da [AFOEM \(1999\)](#) ficamos com o absenteísmo motivado por incapacidade física e psicológica, seja ela motivada por doenças ocupacionais ou por adoecimento sem relação com a atividade profissional.

## 2.2 Aprendizagem Profunda

Hoje, computadores são capazes de entender comandos por voz, reconhecer rostos e expressões faciais, entender letras e símbolos, mesmo que em escrita cursiva, compreender a relação entre eventos passados e presentes e usar essas informações para fazer previsões, são capazes de analisar e fazer diagnósticos com base em exames de imagens, controlar os melhores parâmetros de pacientes internados em UTIs, controlar robôs, apoiar pesquisas científicas básicas etc ([GOODFELLOW et al., 2016](#)).

Toda essa inovação só se tornou possível pela adoção de tecnologias que aproximaram as máquinas de rotinas naturalmente executadas por seres humanos. Para [Goodfellow et al. \(2016\)](#) o verdadeiro desafio para a inteligência artificial é tornar as máquinas capazes de resolver tarefas que são intuitivamente fáceis de serem resolvidas por seres humanos, como o reconhecimento de rostos de pessoas, da voz humana, diferenciar automóveis, pessoas e animais numa imagem, mas que continuam extremamente complexas quando convertidas para o formalismo aceito pelas máquinas.

Logo, a programação de máquinas baseadas em regras rígidas mostrou-se não ser eficiente na execução dessas tarefas, ditas simples. Como solução, esse paradigma da programação formal foi alterado para que as máquinas fossem capazes de aprender pela experiência, que pudessem construir suas regras, para que o mundo fosse entendido como em termos de hierarquias de conceitos relacionados dos mais complexos até chegar-se aos conceitos mais simples. Esse tipo de abordagem evitaria a necessidade da intervenção de um operador humano para especificar todos

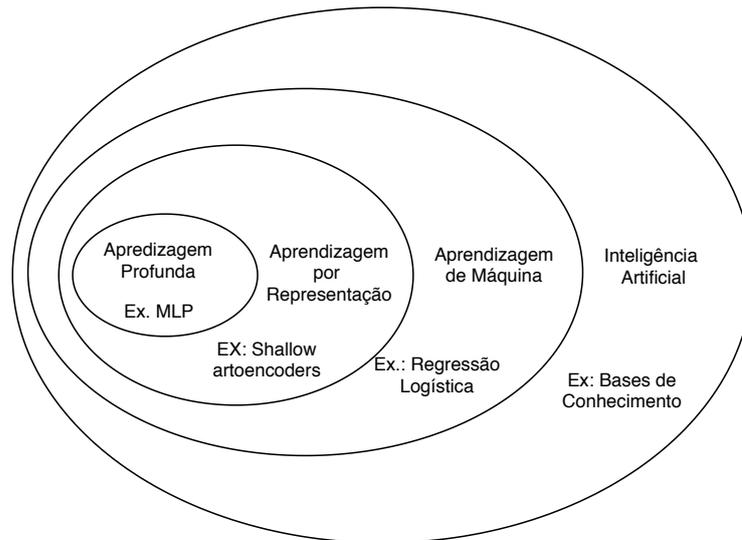


Figura 2 – Um diagrama de *Venn* mostrando como a aprendizagem profunda é um tipo de aprendizado por representação, que é, por sua vez, um tipo de aprendizado de máquina, que é usado para muitas, mas nem todas as abordagens de Inteligência Artificial. Cada seção do diagrama de *Venn* inclui um exemplo de uma tecnologia de IA.

Fonte: [Goodfellow et al. \(2016\)](#)

os conhecimentos que o computador precisa para inferir ou realizar uma tarefa ([GOODFELLOW et al., 2016](#)).

Entretanto, esses conceitos, ainda sim, devem ser representados de forma que as máquinas consigam entendê-los, o que por si só, já é uma tarefa bastante árdua. Para [Goodfellow et al. \(2016\)](#), o meio para superar essas dificuldades seria habilitar os sistemas de Inteligência Artificial (IA) com a capacidade de adquirir seus próprios conhecimentos, por meio da extração de dados brutos obtidos de aplicações da vida real, essa capacidade é denominada aprendizagem de máquina.

Assim, a Inteligência Artificial expandiu-se para abarcar as mais diversas formas pelas quais o conhecimento poderia ser extraído da maneira mais automática possível, como podemos observar na Figura 2 que traz o diagrama de *Venn* com a hierarquia entre as abordagens e exemplos de tecnologias mais utilizadas para esse feito.

Assim, uma das tarefas iniciais de qualquer atividade de implementação de uma tecnologia de IA é definir quais serão as características a serem extraídas e analisadas para a determinação do conjunto de *features* que serão usadas pela máquina para que essa possa inferir os conceitos necessários para a realização da tarefa proposta. Podendo, inclusive utilizar a própria aprendizagem de máquina para realizar essa tarefa de descoberta de forma automatizada, como citamos anteriormente.

A aprendizagem profunda irá atuar nesse problema de representação do conhecimento, permitindo que as características gerais sejam obtidas por meio da análise de características mais simples, descobrindo as relações existentes a fim de construir os conceitos complexos que

denotam o conhecimento a ser representado. Nesse sentido, a própria máquina aprenderia a representação correta para os dados fornecidos a ela como entrada, permitindo fazer uma relação com os dados de saída.

Segundo [Goodfellow et al. \(2016\)](#), o conceito de profundo estaria relacionado ao fato de que a máquina poderia aprender um programa de computador com várias etapas, onde cada camada da representação pode ser considerada como o estado da memória do computador depois de executar outro conjunto de instruções em paralelo. Assim, redes com maior profundidade poderiam executar mais instruções em sequência. Essa profundidade, poderia inclusive, permitir que os resultados obtidos por uma camada de representação pudessem servir de entrada para camadas de representação anteriores.

A medida dessa profundidade poderia ser estabelecida de duas formas. Na primeira, pela mensuração da quantidade de instruções sequenciais necessárias para que o modelo avalie suas saídas, ou seja, qual o caminho mais longo necessário para que, por meio de um fluxograma, cada resultado possível do modelo seja alcançado. Na segunda, a profundidade não estaria relacionada a profundidade do fluxograma do modelo, mas na profundidade do mapa que descreve cada conceito e como eles estão relacionados entre si, podendo ser muito mais profunda que o próprio fluxograma do modelo ([GOODFELLOW et al., 2016](#)).

De forma resumida, a aprendizagem profunda é um tipo específico de aprendizagem de máquina, que representa uma forma de abordagem da Inteligência Artificial. Destaca-se por permitir que sistemas computacionais, baseados na experiência e em dados, aprendam como lidar com problemas do mundo real, permitindo grande flexibilidade no tratamento de conceitos complexos que podem ser tratados como conceitos mais simples hierarquizados e relacionados. Todas essas disciplinas estão relacionadas como podemos observar no diagrama de Venn mostrado na [Figura 2](#).

## 2.3 Aprendizado Supervisionado

O aprendizado supervisionado, também denominado aprendizado com um professor, caracteriza-se pela adoção da analogia de um professor que possui conhecimento sobre o ambiente. Esse conhecimento é representado por um conjunto de exemplos para os quais, para cada *entrada* há uma respectiva *saída* ótima. Nessa estrutura, o modelo que irá ser *ensinado* não possui o conhecimento sobre o ambiente. Nesse contexto, se tanto o professor e o modelo a ser treinado forem expostos a exemplos extraídos desse ambiente, o professor, devido ao seu conhecimento prévio, poderá fornecer ao modelo a resposta ideal ao exemplo fornecido como entrada ([HAYKIN, 2001](#)).

Nesse cenário, para um determinado neurônio  $k$  no tempo de interação  $t$  o modelo a ser treinado tem seus parâmetros ajustados pela relação entre o vetor de treinamento  $x(t)$  e o erro  $e_k(t)$ . Para isso, o erro  $e_k(t)$  representa a diferença entre o valor esperado  $d_k(t)$  e a resposta

efetivamente fornecida pela arquitetura  $y_k(t)$  no tempo de iteração  $t$  como na Equação 2.1 (HAYKIN, 2001).

$$e_k(t) = d_k(t) - y_k(t) \quad (2.1)$$

O erro  $e_k(t)$  aciona um sistema de correção que é realizada a cada iteração, passo a passo, a fim de que a arquitetura simule as ações do professor, aproximando os valores obtidos como saída  $y_k(t)$  daquele informado pelo professor  $d_k(t)$ . O comportamento esperado para a arquitetura é que essa chegue ao ponto de não haver mais a necessidade do professor, o que representaria a transferência do conhecimento do professor para a arquitetura. Esse objetivo é obtido pela minimização de uma função de custo ou índice de desempenho  $\varepsilon(t)$ , conforme Equação 2.2 (HAYKIN, 2001).

$$\varepsilon(t) = \frac{1}{2}e_k^2(t) \quad (2.2)$$

Essa forma de aprendizagem supervisionada pode ser também denominada aprendizagem por correção de erro. Na Figura 3, percebemos que se caracteriza por ser um sistema realimentado e de ciclo fechado em que o ambiente compartilhado entre a arquitetura e o professor e que serve de subsídio ao treinamento não pertence ao laço.

Para aferir o desempenho desse tipo de arquitetura quanto ao treinamento realizado, adota-se a *taxa de erro quadrático*, ou outra medida de erro, sobre a amostra de treinamento em função dos parâmetros livres do sistema. A visualização do resultado dessa função é uma *superfície multidimensional de desempenho de erro* ou *superfície de erro*, que tem os parâmetros livres como coordenadas (HAYKIN, 2001).

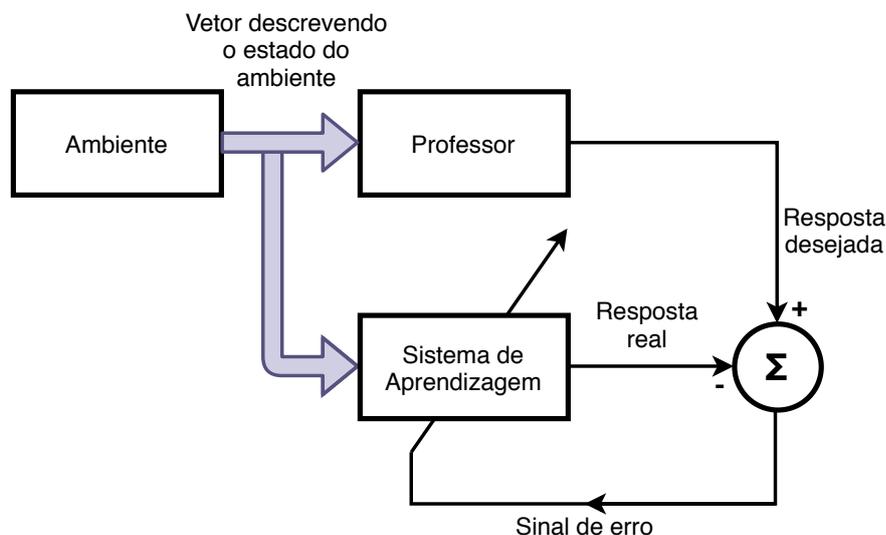


Figura 3 – Diagrama em blocos da aprendizagem com um professor  
Fonte: Haykin (2001)

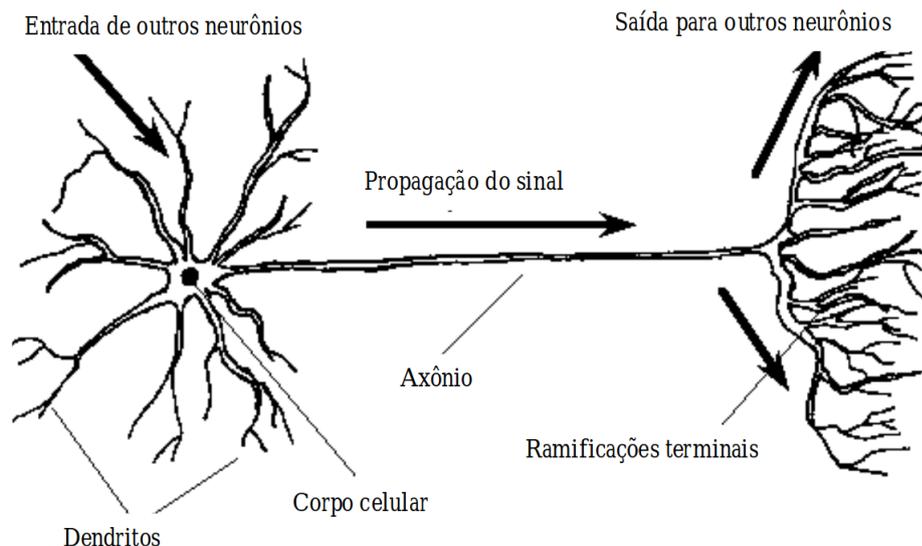


Figura 4 – Célula neural biológica com a sequência de propagação do sinal  
Fonte: Silva (1998)

Nessa superfície todas as operações realizadas sob a supervisão do professor são projetadas por meio de um ponto. O aprendizado nesse sistema e a consequente melhora no desempenho da arquitetura é representado pela movimentação desse ponto para uma localização mínima na *superfície de erro*. Esse movimento em direção a localização mínima sobre a *superfície de erro* é orientado pelo vetor que aponta na direção da descida mais íngreme, também denominado *vetor gradiente* (HAYKIN, 2001).

## 2.4 Redes Neurais Artificiais

As Redes Neurais Artificiais surgiram da inspiração de trazer aos computadores a eficiência com que o cérebro humano processa informações, de forma não-linear, complexa e em paralelo. Para tanto, o cérebro se utiliza de estruturas denominadas neurônios como os representados na Figura 4, capazes de se organizar de modo a realizar processamentos simultâneos, se aproveitando de diversos receptores sensoriais que o habilitam a reconhecer padrões, como rostos familiares numa multidão, reconhecer sons específicos, o timbre de voz de uma criança etc (HAYKIN, 2001).

Assim, uma rede neural é a modelagem dessa habilidade do cérebro humano, que permite que as máquinas aprendam com a experiência a que são submetidas por meio de entradas que são fornecidas pelo ambiente, como no desenvolvimento do cérebro de uma criança. Como nos cérebros humanos e de outros animais, esse aprendizado é realizado por pequenas unidades de processamento (Figura 5), que são interligadas umas às outras de modo a simular as conexões existentes entre os neurônios biológicos.

Para Haykin (2001) uma rede neural é um processador maciçamente e paralelamente distribuído constituído de unidades de processamento simples, que têm a propensão natural para

armazenar conhecimento experimental e torná-lo disponível para uso.

Outra característica citada pelo autor para definir uma rede neural artificial é que o conhecimento da rede é adquirido a partir do ambiente, esse processo é denominado *processo de aprendizagem*. Outra estaria relacionada a força que conecta os neurônios artificiais, que seriam denominadas, por referência aos neurônios biológicos de *pesos sinápticos* e seriam utilizados para armazenar o conhecimento adquirido pela rede (HAYKIN, 2001).

Para que a rede aprenda esses *pesos sinápticos* precisam ser modificados, utilizando-se para isso de algoritmos denominados *algoritmos de aprendizagem* que fazem a atualização dos valores existentes nesses pesos de forma ordenada para que a rede atinja ao objetivo a que foi proposto, como veremos nas próximas seções.

### 2.4.1 Estrutura de um neurônio artificial

As redes neurais, como dito, são formadas por unidades mais simples que se conectam, denominadas *neurônios artificiais*, ou simplesmente *neurônios*. Na Figura 5 mostramos um modelo de neurônio  $k$  não-linear, composto por algumas partes básicas, dentre elas podemos destacar a entrada no neurônio dos estímulos  $\{x_1, x_2, \dots, x_j\}$  por meio das *sinapses*. Para cada sinapse há um peso associado correspondente  $w_{kj}$ , em que o primeiro índice  $k$  indica a qual neurônio aquele peso sináptico pertence e o segundo  $j$  a qual terminal de entrada do neurônio está associado (HAYKIN, 2001).

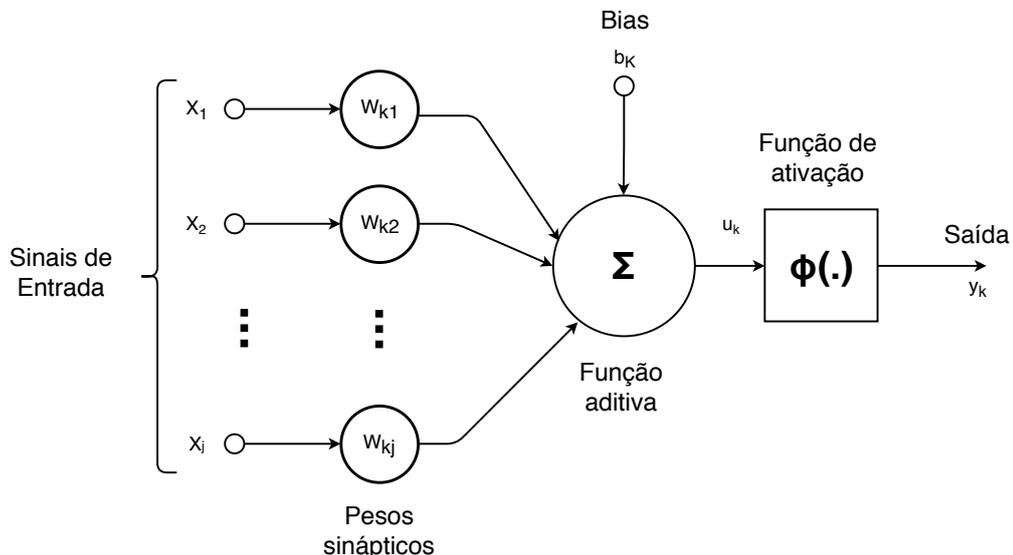


Figura 5 – Modelo não-linear de um neurônio artificial recebendo as entradas  $x_j$ , sendo processado pelo neurônio  $k$ , com uma função de ativação  $\varphi$  e tendo como saída  $y_k$

Fonte: Haykin (2001)

$$u_k = \sum_{j=1}^m w_{kj} x_j \quad (2.3)$$

$$y_k = \varphi(u_k + b_k) \quad (2.4)$$

Após os pesos sinápticos, temos uma função somadora  $\Sigma$  ponderada pelos pesos sinápticos. Em seguida, aparece a *função de ativação*  $\varphi$  adotada pela arquitetura do neurônio. Essa tem a finalidade de restringir quais são os valores de saída possíveis para o neurônio, além disso, nesse modelo existe um *bias*  $b_k$  que serve para aumentar ou diminuir a entrada líquida da função de ativação. Mais precisamente, a saída do neurônio será obtida pelo resultado da Equação 2.3 e da Equação 2.4 aplicadas sobre os sinais de entrada  $\{x_1, x_2, \dots, x_j\}$  (HAYKIN, 2001). Falaremos sobre cada atividade descrita nessa seção e seus respectivos algoritmos de implantação mais comuns nas seções a seguir.

## 2.4.2 Funções de ativação

A função de ativação de um neurônio artificial como o esquematizado na Seção 2.4.1 por  $\varphi$  é aquela responsável por definir o intervalo aceito como saída do neurônio após estimulado. Podem ser empregadas diversos tipos de funções de ativação, de acordo com a necessidade da aplicação a que se destinará a rede neural.

Descreveremos alguns desses tipos a seguir, pois serão utilizados em nossos experimentos. E como observaremos em algumas arquiteturas de redes neurais, elas serão usados em conjunto para a obtenção da saída esperada por alguns dos tipos de redes neurais empregadas.

### 2.4.2.1 Função de ativação do tipo retificadora

As funções de ativação do tipo retificadoras são bastante semelhantes as funções lineares. A única diferença entre as funções é que a retificadora, quando recebe valores negativos, retorna como saída o valor zero, ou seja, a imagem dessas funções é a metade daquele de uma função linear propriamente dita. Nessa condição, uma função de ativação retificadora sempre retorna valores positivos e também diretamente proporcionais aos valores recebidos como entrada. Sua expressão é determinada pela Equação 2.5 e tem como gráfico típico o mostrado na Figura 6a (GOODFELLOW et al., 2016).

$$g(z) = \max\{0, z\} \quad (2.5)$$

Sua maior vantagem é ter a derivada sempre 1 ou 0, para  $z \neq 0$ . Uma desvantagem desse tipo de função de ativação retificada é que elas não podem aprender através de métodos baseados em gradientes em exemplos para os quais sua ativação é zero. Uma forma de compensar essa

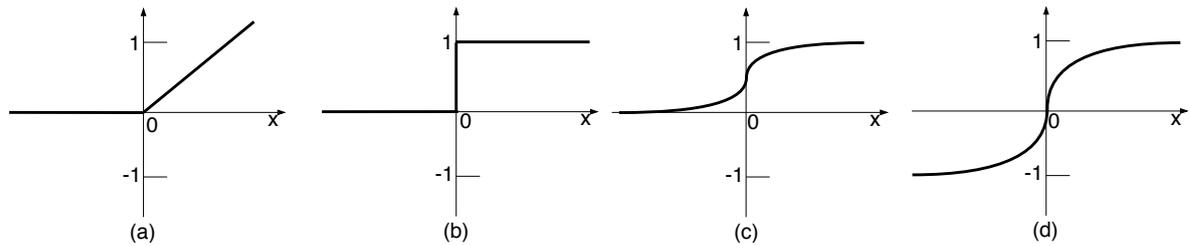


Figura 6 – Adaptação dos gráficos típicos das funções de ativação relacionadas por Michael (2005) como as mais bem sucedidas em aplicações práticas: (a) retificadora; (b) função por limiar; (c) função sigmoide; (d) função tangente.

Fonte: Michael (2005)

dificuldade é atribuir sempre aos valores recebidos, valores pequenos de *bias* evitando-se que a saída seja desativada (GOODFELLOW et al., 2016).

#### 2.4.2.2 Função de ativação por limiar

A função de ativação por limiar, também conhecida como *função de Heaviside* converte a saída de um neurônio para um intervalo discreto pré-estabelecido, geralmente 0 ou 1, como descrito na Equação 2.6 e representada na Figura 6b (HAYKIN, 2001).

$$g(z) = \begin{cases} 1 & \text{se } z \geq 0 \\ 0 & \text{se } z < 0 \end{cases} \quad (2.6)$$

Ela representa a aplicação do modelo *McCulloch-Pits* estabelecido no trabalho pioneiro desses autores realizado em 1943, em que o modelo do neurônio tinha como paradigma o *tudo-ou-nada*, saída 1 para ativado e saída 0 quando não ativado (HAYKIN, 2001).

Podem ainda ser empregadas funções de ativação por limiar compostas, em que seriam estabelecidos diversos níveis de ativação de acordo com os valores retornados pela função de adição do neurônio  $u_k$  (HAYKIN, 2001).

#### 2.4.2.3 Função de ativação do tipo sigmoide

A função sigmoide apresenta um gráfico em forma de *S* e segundo Haykin (2001) é a mais comum na construção de redes neurais artificiais. É uma função crescente que permite um balanceamento adequado entre comportamentos lineares e não-lineares. Como exemplo dessa função temos a função logística descrita na Equação 2.7 e representada na Figura 6c.

$$g(z) = \varphi(u) = \frac{1}{1 + e^{(-au)}} \quad (2.7)$$

Na Equação 2.7,  $a$  representa o coeficiente de inclinação da função sigmoide, se esse parâmetro se aproximar do limite, tendendo ao infinito, teremos uma função semelhante a limiar. Já  $u$  representa o resultado da função de adição do neurônio. Um diferencial desse tipo de função

de ativação é que enquanto a função de ativação por limiar assume os valores discretos 0 ou 1, a função de ativação sigmoideal assume valores contínuos nesse intervalo (HAYKIN, 2001).

Para Goodfellow et al. (2016) as funções de ativação do tipo sigmoide devem ser evitadas quando utilizadas em camadas ocultas de redes neurais do tipo *feedforward*, pois como visto na Equação 2.7, são muito sensíveis a valores altos positivos de entrada para  $u$  ou quando esse valor é muito alto, mas negativo, aproximando-se muito de 0, o que dificultaria o aprendizado da rede. Sendo sua utilização mais recomendada nas camadas de saída da arquitetura.

#### 2.4.2.4 Função de ativação do tipo tangente hiperbólica

Segundo Goodfellow et al. (2016), as funções de ativação do tipo tangente hiperbólica devem ser preferencialmente escolhidas quando se pensar na utilização de uma função de ativação do tipo sigmoide. Segundo o autor, a função de ativação baseada em tangente hiperbólica normalmente apresenta um desempenho melhor do que uma função de ativação do tipo sigmoideal logística. Como exemplo, cita o caso da função identidade obtida quando o neurônio tem como entrada o valor 0. Enquanto para  $\tanh(0) = 0$ , para uma função  $\sigma(0) = \frac{1}{2}$ . Contudo, para ele, apesar dos problemas inerentes ao risco de saturação das funções de ativação do tipo sigmoide, essas são bem vindas em outras estruturas de redes como as recorrentes.

$$g(z) = \tanh(z) = \frac{e^z - e^{-z}}{e^z + e^{-z}} \quad (2.8)$$

As funções de ativação do tipo tangente hiperbólicas assumem valores contínuos no intervalo entre  $-1$  e  $+1$ , como podemos depreender pelos valores retornados pela Equação 2.8 e representada na Figura 6d (GOODFELLOW et al., 2016).

#### 2.4.3 Função de custo

Um dos importantes aspectos a serem definidos quando da construção de uma rede neural é a escolha da função de custo. Essas guardam semelhanças com as adotadas por outras metodologias, como os modelos lineares. Esse tipo de função é usada para avaliar a diferença obtida pelo resultado da rede e aquele esperado.

Em muitos modelos a função de custo representa a distribuição da probabilidade de  $y$  a partir de  $x$  por  $p(y|x; \Theta)$  usando o princípio da máxima semelhança, ou seja, é usado a *cross-entropy* entre os dados de treino e a predição do modelo como função de custo (GOODFELLOW et al., 2016).

Uma função com essa característica pode ser vista na Equação 2.9 e também na Equação 2.2:

$$J(\Theta) = -\mathbb{E}_{x,y \sim \hat{p}_{data}} \log p_{model}(y|x) \quad (2.9)$$

Para [Goodfellow et al. \(2016\)](#), outro importante cuidado que se deve ter no projeto de redes neurais é para que as funções de custo tenham um gradiente suficientemente grande que permita que o algoritmo de aprendizado o tenha como guia. Para isso, também deve-se ter cuidado com a função de ativação escolhida que pode gerar um gradiente aproximadamente plano, o que dificultaria também o processo de aprendizado. Geralmente essa situação ocorre pela saturação de funções baseadas em operações exponenciais que recebem como entrada valores negativos muito altos, por isso o emprego da probabilidade negativa ajudaria na minimização desse problema na função de cálculo do custo, como observamos na Equação 2.9.

#### 2.4.4 Algoritmo *Back-propagation*

Quando uma rede artificial do tipo *feedforward* recebe como entrada um valor  $x$  e produz como saída um valor  $y'$ , isso significa que a informação percorreu a rede, desde sua entrada, passando por suas camadas ocultas até produzir o valor  $y'$ . Esse processo é denominado *forward propagation*. Essa atividade acontece durante o treinamento produzindo um custo escalar  $J(\Theta)$ . É por meio do algoritmo de retropropagação (*back-propagation*), que esse custo do treinamento é irradiado para trás, pela rede, permitindo que se calcule o gradiente ([GOODFELLOW et al., 2016](#)).

Segundo [Goodfellow et al. \(2016\)](#) existe em alguns casos uma confusão com a definição do que faz um algoritmo do tipo *back-propagation* que é muitas vezes associado a todo o processo de aprendizado de uma rede neural. Entretanto, ele é responsável apenas pela construção do gradiente, onde outro algoritmo, como o método do gradiente estocástico, será o responsável pelo aprendizado propriamente dito.

Para o cálculo do gradiente de  $f(x, y)$  para uma função arbitrária  $f$ , onde  $x$  é um conjunto de variáveis cujas derivadas são desejadas e  $y$  representa também um conjunto adicional de variáveis de entrada da função, mas para as quais a derivada não é desejada. A principal função do algoritmo *back-propagation* é o cálculo do gradiente da função de custo adotada em relação aos parâmetros fornecidos, como em  $\nabla_{\Theta} J(\Theta)$  ([GOODFELLOW et al., 2016](#)).

Esse gradiente é obtido por meio do *Cálculo da Regra de Cadeia*, que consiste na obtenção das derivadas das funções formadas pela composição de outras funções cujas derivações são conhecidas. Assim, o algoritmo *back-propagation* seria o responsável pela aplicação da regra da cadeia em uma ordem específica e de forma eficiente ([GOODFELLOW et al., 2016](#)).

Assim, se temos  $x$  como um número real e  $f$  e  $g$  como funções tais que  $f : \mathbb{R} \rightarrow \mathbb{R}$  e  $g : \mathbb{R} \rightarrow \mathbb{R}$ , se:

$$y = g(x) \tag{2.10}$$

$$z = f(g(x)) = f(y) \tag{2.11}$$

Pela regra da cadeia para as derivadas teríamos:

$$\frac{dz}{dx} = \frac{dz}{dy} \frac{dy}{dx} \quad (2.12)$$

Seguindo, se tivermos  $x \in \mathbb{R}^m$  e  $y \in \mathbb{R}^n$ , a função  $g$  definida por  $g : \mathbb{R}^m \rightarrow \mathbb{R}^n$ , a função  $f$  definida por  $f : \mathbb{R}^n \rightarrow \mathbb{R}$  e se  $y = g(x)$  e  $z = f(y)$ , então teríamos por analogia:

$$\frac{\partial z}{\partial x_i} = \sum_j \frac{\partial z}{\partial y_j} \frac{\partial y_j}{\partial x_i} \quad (2.13)$$

Convertendo para notação de vetores, poderíamos escrever a Equação 2.13 da seguinte forma:

$$\nabla_x z = \left( \frac{\partial y}{\partial x} \right)^\top \nabla_y z \quad (2.14)$$

Sendo o gradiente obtido pela multiplicação da matriz Jacobiana  $\frac{\partial y}{\partial x}$  pelo gradiente  $\nabla_y z$  e calculado para cada operação no grafo da rede (GOODFELLOW et al., 2016).

## 2.4.5 Algoritmo de otimização

Existem diversos tipos de algoritmos de otimização, os mais comuns utilizados em arquiteturas do tipo *feedforward* são aqueles baseados no gradiente. Como vimos na Seção 2.4.4, há a distinção entre o algoritmo responsável pela montagem do gradiente e aquele que irá fazer a otimização percorrendo esse gradiente em busca de seus valores mínimos ou máximos, como veremos adiante.

A otimização é realizada a cada passo da iteração na fase denominada *treinamento*, no entanto, para que haja um aproveitamento dos dados recebidos anteriormente, alguns algoritmos adotam funções do tipo estocástica. Em nosso trabalho, usamos algoritmos desse tipo, que se fundamentam no aprendizado por meio da otimização baseada no gradiente. Desse modo, para o nosso trabalho adotamos o Adam (KINGMA; BA, 2014).

Segundo seu autor, esse algoritmo tem como vantagem diante de outros algoritmos de otimização baseados em gradiente, ser montado sobre estimativas adaptativas de momentos de baixa ordem. O algoritmo seria também simples de implementar, computacionalmente eficiente, consumiria pouca memória e seria adequado à aplicação em problemas grandes em termos de dados e/ou em número de parâmetros. Além disso, funcionaria inclusive em gradientes ruidosos ou esparsos e requereria poucos ajustes nos seus hiperparâmetros (KINGMA; BA, 2014).

Os principais parâmetros para o algoritmo do tipo Adam e seus respectivos valores padrões seriam o  $\alpha = 0.0001$ , o  $\beta_1 = 0.9$ , o  $\beta_2 = 0.999$  e  $\epsilon = 10^{-8}$ . O  $\alpha$  representaria o tamanho do passo do algoritmo sobre o gradiente, o  $\beta_1$  e  $\beta_2$  são as taxas de decaimento

exponencial para as estimativas de momento e o  $\epsilon$  é usado como parâmetro de regularização na fase de atualização dos parâmetros (KINGMA; BA, 2014).

Para o cálculo da atualização dos parâmetros, vários passos são realizados, como no algoritmo que o autor explicita em seu trabalho, que em resumo seriam a inicialização das variáveis  $\theta_0$  que são os vetores iniciais; setar o valor de  $f(\theta)$  que seria a função objetivo com parâmetros  $\theta$ ; inicialização do  $m_0$  que seria o primeiro vetor de momento; a inicialização de  $v_0$  que seria o segundo vetor de momento e  $t$  que é o tempo de execução do algoritmo.

Feitos esses passos, seguiria-se aplicando sucessivamente as Equações 2.15 até a 2.21, até que o algoritmo encontre o conjunto de parâmetros  $\theta_t$  ótimos para o gradiente informado.

$$t \leftarrow t + 1 \quad (2.15)$$

$$g_t \leftarrow \nabla_{\theta} f_t(\theta_{t-1}) \quad (2.16)$$

$$m_t \leftarrow \beta_1 \cdot m_{t-1} + (1 - \beta_1) \cdot g_t \quad (2.17)$$

$$v_t \leftarrow \beta_2 \cdot v_{t-1} + (1 - \beta_2) \cdot g_t^2 \quad (2.18)$$

$$\hat{m}_t \leftarrow \frac{m_t}{1 - \beta_1^t} \quad (2.19)$$

$$\hat{v}_t \leftarrow \frac{v_t}{1 - \beta_2^t} \quad (2.20)$$

$$\theta_t \leftarrow \theta_{t-1} - \frac{\alpha \cdot \hat{m}_t}{\sqrt{\hat{v}_t} + \epsilon} \quad (2.21)$$

Esse algoritmo é processado a cada passo da iteração no tempo  $t$  até que a função convirja para um mínimo no gradiente formado pelo algoritmo *Back-propagation*.

Esse algoritmo foi o adotado para ser utilizado no trabalho pois mostrou-se realmente eficiente quando aplicado aos dados do estudo, além de ter sido comparado com outros algoritmos de otimização como *RMSProp*, *SGD*, *Adamax*, *Nadam*, *Adagrad*, *Adadelta* por ocasião de outras disciplinas do curso.

## 2.5 Arquiteturas utilizadas

Foram selecionadas algumas classes de aprendizagem profunda de diferentes naturezas. Optamos pelas arquiteturas do tipo *Multilayer Perceptron*, *Recurrent Neural Network* e

*Long Short-Term Memory*. Para avaliarmos se essas arquiteturas poderiam alcançar resultados superiores aos resultados obtidos por outras técnicas de Aprendizagem de Máquina, adotamos a arquitetura do tipo Máquina de Vetores de Suporte como referência, pelo seu sucesso em aplicações de classificação e sua capacidade em lidar com um elevado número de dimensões.

Para o trabalho, foi adotado o aprendizado supervisionado, ou seja, para cada entrada fornecida, há uma saída correspondente que representa a classificação do exemplo. Isto permite que as arquiteturas infiram quais são as características comuns a cada item a ser classificado, baseando-se na técnica par de entrada-saída, permitindo que a arquitetura possa classificar um novo exemplo com base na *experiência* adquirida durante o treinamento. Sendo assim, seguem as principais características das arquiteturas utilizadas no trabalho.

### 2.5.1 Arquitetura do tipo *Multilayer Perceptron*

As redes neurais artificiais foram inspiradas na tentativa de simulação dos sistemas neurais encontrados nos seres biológicos, que possuem como unidade básica o neurônio, que são células nervosas ligadas umas às outras através de fios de fibra denominados axônios. Esses últimos são os responsáveis pelo transporte da informação, representada pelos impulsos nervosos. Um axônio é ligado a outro neurônio por meio de estruturas denominadas dendritos, e esse ponto de contato é denominado sinapse (SILVA, 1998).

Buscando transportar essa estrutura para as redes neurais artificiais, foi criada uma estrutura análoga denominada perceptron, que seria o equivalente a um neurônio artificial. Esses perceptrons podem ser arranjados de diversos modos, a fim de elevar suas capacidades de classificação. Uma dessas formas é a *Multilayer Perceptron* que possui uma estrutura mais complexa que o simples modelo perceptron. Essa estrutura, geralmente, se utiliza de uma função de correção de erro denominada *back-propagation* que permite a atualização dos pesos atribuídos a cada nodo da rede a fim de encontrar aqueles que minimizem a taxa de erro de classificação, como vimos na Seção 2.4.4 (PANG-NING; STEINBACH; KUMAR, 2009).

Assim, as redes neurais artificiais do tipo *Multilayer Perceptron* (MLP) têm se mostrado bastante eficientes na resolução de problemas não linearmente separáveis. Isso só foi possível com o advento de estruturas mais complexas, onde foram adicionadas camadas intermediárias às estruturas anteriores de camada única, ou perceptron. A estrutura característica para esse tipo de arquitetura de rede é aquela em que todos os neurônios da camada anterior estão densamente conectados aos neurônios da camada seguinte (FACELI et al., 2011). Um exemplo desse modelo de arquitetura de rede neural artificial pode ser visto na Figura 7.

Assim, esse tipo de rede neural artificial denominada *Multilayer Perceptron* (MLP) é representado por um conjunto de unidades sensoriais que funcionam como a camada de entrada da rede (*input layer*). Essa camada receberá os “estímulos” representados pelos dados de entrada (*input*). Após essa camada, estão localizadas uma ou mais camadas densamente conectadas

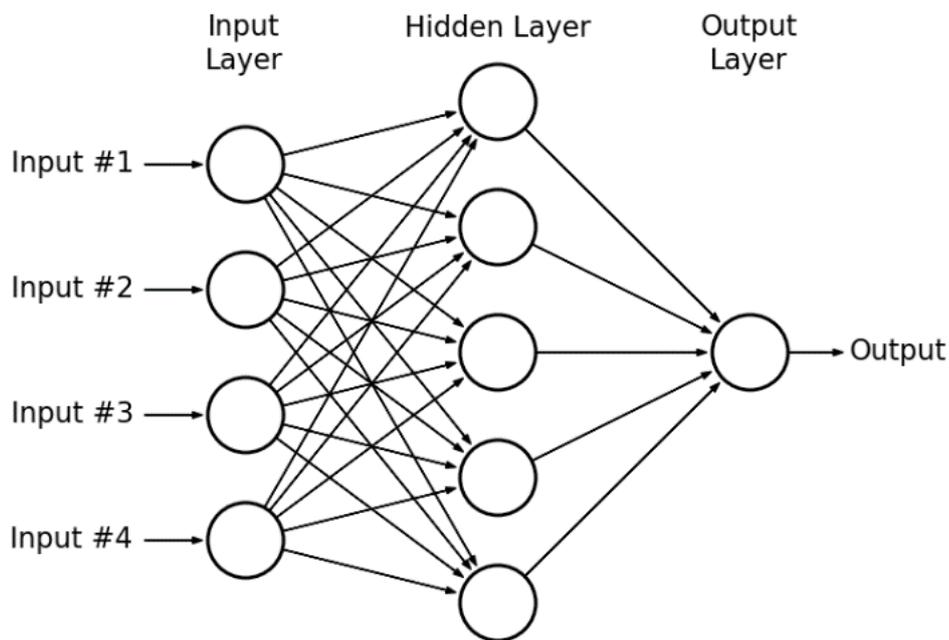


Figura 7 – Um exemplo hipotético de uma rede *Multilayer Perceptron*  
 Fonte: Hassan et al. (2015)

denominadas camadas ocultas (*hidden layer*), seguidas por mais uma camada, também ligada densamente por nós computacionais, denominada camada de saída (*output layer*), responsável pelo resultado final da classificação da amostra (*output*) (FACELI et al., 2011).

## 2.5.2 Arquitetura do tipo *Recurrent Neural Network*

As arquiteturas do tipo *Recurrent Neural Network* (RNN) são modelos de redes capazes de tratar dados sequenciais. Esses dados estão presentes em diversos problemas modernos, como reconhecimento de linguagem, reconhecimento de escritas cursivas e em problemas em que dados de séries temporais são dados como entrada da rede (CHO et al., 2014).

São caracterizadas por um estado oculto  $h$  e uma saída opcional  $y$  que recebe uma sequência variável de entradas  $x = (x_1, \dots, x_t)$ , e a cada tempo  $t$ , o estado oculto  $h_{<t>}$  é atualizado por uma função de ativação  $f$  como na Equação 2.22 (CHO et al., 2014).

$$h_{(t)} = f(h_{(t-1)}, x_t) \quad (2.22)$$

Logo, esse tipo de arquitetura consegue tratar problemas que guardam relação com dados anteriormente recebidos porque em seus nós computacionais há a presença de *loops*, o que permite que a informação seja guardada e transmitida para as próximas iterações. Portanto, esse tipo de arquitetura possui uma memória.

No esquema mostrado na Figura 8, a entrada da célula  $A$  no tempo  $t$  é representada por  $X_t$ , função de ativação é a  $\tanh$ . A saída da célula  $h_t$  é representada pelo valor obtido pela

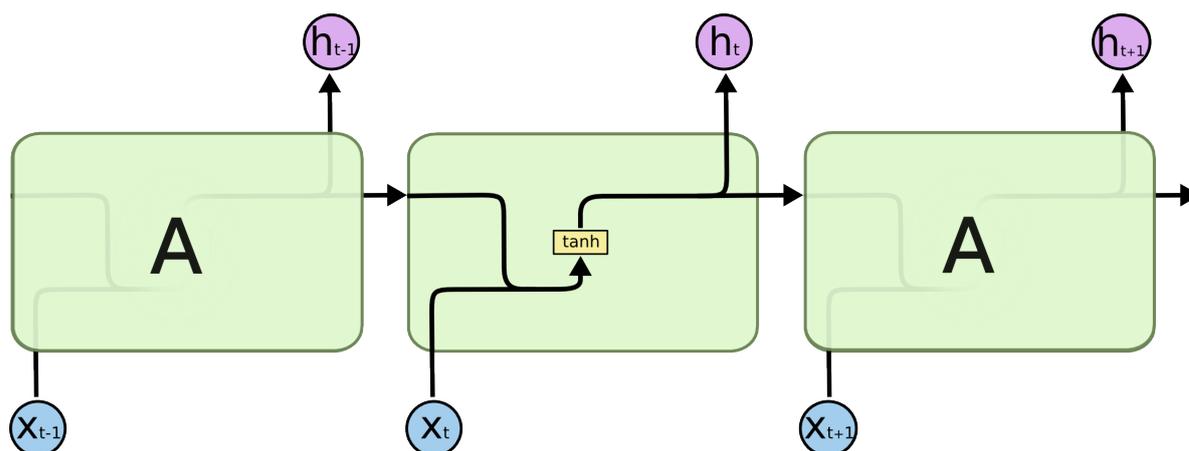


Figura 8 – Estrutura esquematizada de uma célula RNN e suas conexões recorrentes nos tempos  $t - 1$ ,  $t$  e  $t + 1$

Fonte: Olah (2015)

concatenação de  $h_{t-1}$  e  $x_t$  aplicada à função de ativação é a tangente hiperbólica, descrita na Seção 2.4.2.4 e na Figura 8 denominada *tanh*. O estado da célula  $A$  é atualizado nesse momento, representado pelo valor de  $h_t$  que é novamente usado como entrada da célula  $A$  no tempo  $t + 1$  (OLAH, 2015).

### 2.5.3 Arquitetura do tipo *Long Short-term Memory*

Assim como as RNN, as arquiteturas de redes do tipo *Long Short-term Memory* (LSTM) se diferenciam das arquiteturas do tipo MLP por terem sido desenvolvidas para tratar dados que guardam relações com outros dados anteriormente recebidos pela rede em relação ao tempo, que podem guardar dependências, inclusive, com os dados seguintes, como numa memória de curto e longo prazo. Essa arquitetura se enquadra no grupo de redes neurais recorrentes e funciona como caminhos que os dados percorrem, passando por portões que permitem gerenciar quais informações são “armazenadas” e quais informações serão “esquecidas”, podendo atualizar os pesos das conexões a cada iteração (GOODFELLOW et al., 2016). Sendo essa característica o que as diferencia em relação a outros tipos de RNN.

Diferencia-se também das MLP porque em vez da utilização de somente neurônios artificiais como nas redes do tipo Multilayer Perceptron, esses são agrupados em “células” que são conectadas recursivamente entre si. Essas células recebem os dados que podem ou não alterar o estado da célula, dependendo do controle executado pelo “portão de entrada”  $x_t$ , que utiliza uma função sigmoide para determinar essa condição. Se esse estado for preservado, existe na célula um *loop* interno de controle de estado que é gerenciado pela ação de um “portão de esquecimento”. A saída da célula também é controlada por um “portão de saída”  $h$  que pode, inclusive, desligar a saída da célula quando necessário (GOODFELLOW et al., 2016).

Um esquema da estrutura e funcionamento desse tipo de arquitetura é vista na Figura 9. Nele podemos perceber que, como no esquema de uma célula para a arquitetura RNN mostrado

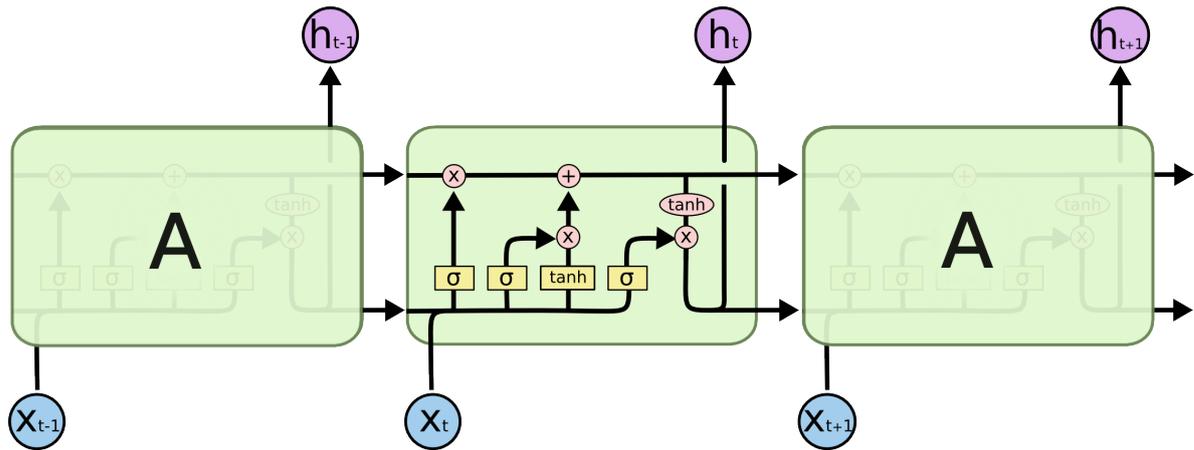


Figura 9 – Estrutura esquematizada de uma célula LSTM e suas conexões recorrentes nos tempos  $t - 1$ ,  $t$  e  $t + 1$

Fonte: Olah (2015)

na Figura 8, a célula da arquitetura LSTM denominada *A*, recebe no tempo  $t$  uma entrada  $X_t$  e produz uma saída  $h_t$ . Contudo, no seu interior, essa célula não possui uma, mas quatro camadas. No esquema da Figura 9 identificamos que cada retângulo em amarelo representa uma camada, os círculos em rosa representam quais operações são realizadas sobre os vetores recebidos, a união de linhas são concatenações de vetores e as bifurcações representam cópias desses vetores (OLAH, 2015).

Além disso, esse tipo de célula possui três estruturas denominadas *portões*, compostas por *perceptrons* que possuem como função de ativação a sigmoide  $\sigma$ , descrita na Seção 2.4.2.3. Após os portões segue-se um operador de multiplicação. Esses *portões* são os responsáveis pela alteração ou conservação do estado da célula. Da esquerda para a direita, o primeiro portão é denominado *portão de esquecimento*, o segundo portão é denominado *portão de entrada* e funciona junto com uma função  $\tanh$  (Seção 2.4.2.4) de ativação para realizar a atualização do estado da célula  $C_t$ . Após essa operação, chegamos ao momento em que é decidido o que será gerado como saída da célula, isso é realizado pelo *portão de saída*, que recebe o estado  $C_t$  da célula e o multiplica pela saída obtida pela função  $\sigma$  do portão de saída. Esse resultado  $h_t$  irá ser concatenado com o próximo valor recebido pela célula em  $X_{t+1}$ , iniciando o ciclo (OLAH, 2015).

Logo, essa arquitetura de rede possui a capacidade de acumular informação, como evidências de uma característica ou categoria procurada e que esteja distribuída ao longo dos dados temporalizados de longa ou curta duração (GOODFELLOW et al., 2016).

#### 2.5.4 Máquina de Vetores de Suporte

A técnica de Aprendizagem de Máquina denominada Máquina de Vetores de Suporte (SVM), tem se mostrado muito eficiente quando usada em problemas de classificação. A configuração de seus hiperparâmetros representa fator preponderante para a eficiência dessa técnica e

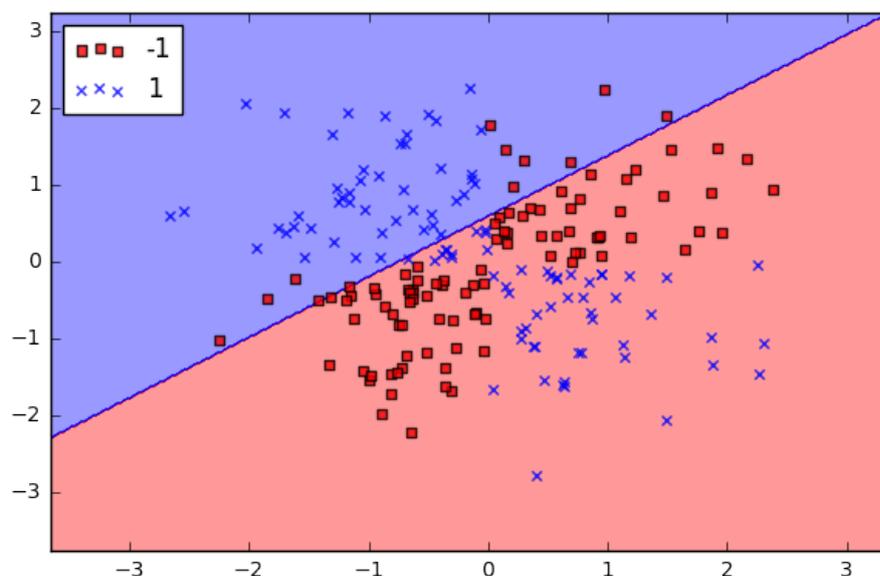


Figura 10 – Exemplo de aplicação de SVM linear para classificação dos elementos de um conjunto de treinamento

Fonte: [Albon \(2017\)](#)

tem sido objeto de diversas pesquisas ([HORVÁTH; MANTOVANI; CARVALHO, 2017](#)).

As SVM se utilizam da teoria da aprendizagem estatística e têm sido utilizadas com sucesso em diversas aplicações práticas. Outra característica importante dessas técnicas é serem capazes de atender a problemas que possuem alta dimensionalidade. Para a classificação dos dados submetidos a essa técnica são fornecidos exemplos de treinamento que são denominados vetores de suporte que servirão de base para a determinação de hiperplanos de separação entre os exemplos de cada classe ([PANG-NING; STEINBACH; KUMAR, 2009](#)). Um exemplo de aplicação de uma SVM linear pode ser vista na Figura 10.

Com uma SVM busca-se um hiperplano que permita obter uma margem de separação ótima, tal que, quando novos dados forem submetidos à SVM, essa seja capaz de classificá-los corretamente em suas respectivas classes, minimizando a taxa de erro de generalização ([PANG-NING; STEINBACH; KUMAR, 2009](#)).

Para nosso trabalho, adotamos uma SVM não linear, haja vista a natureza dos dados não serem, *a priori*, linearmente separáveis. Nesse tipo de abordagem, os dados fornecidos como exemplos para o treinamento da arquitetura são convertidos de suas coordenadas originais  $x$  para coordenadas transformadas  $\Phi(x)$ , onde, nesse novo espaço, poderiam ser separadas.

Usamos uma SVM baseada em *Kernel* do tipo Função de Base Radial (RBF), como na Equação 2.23, que possui como principais parâmetros a serem configurados o  $\gamma$  e o  $C$ . O  $\gamma$  é o hiperparâmetro da arquitetura que define o quanto um único exemplo de treinamento influencia na determinação das regiões de separação. Para isso, um valor baixo para  $\gamma$  significaria uma influência ampla, enquanto, um  $\gamma$  com valores altos, significaria próximos. Pode ser visto como

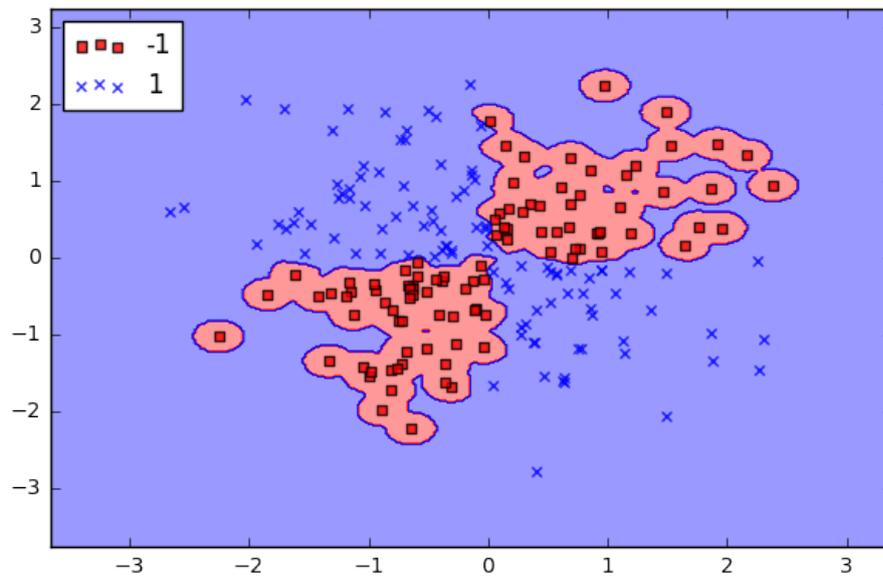


Figura 11 – Exemplo de aplicação de SVM com  $Kernel_{RBF}$  para classificação dos elementos de um conjunto de treinamento

Fonte: [Albon \(2017\)](#)

o inverso do raio da influência das amostras selecionadas pelo modelo como vetores de suporte ([PEDREGOSA et al., 2011](#)).

$$K(x, x') = \sum_i^N \alpha_i y_i \exp\left(-\frac{\|x - x'\|^2}{2\sigma^2}\right) + b \quad (2.23)$$

Já o hiperparâmetro  $C$  representa um opositor à maximização da margem da função de decisão, isso significa que, para valores altos para  $C$ , a SVM aceitará uma margem menor de separação na função de decisão, já para valores menores de  $C$ , a margem de separação será aumentada, sob o risco de redução da precisão do treinamento. Assim, o  $C$  funcionaria como um hiperparâmetro de regularização do SVM ([PEDREGOSA et al., 2011](#)).

Podemos observar uma diferenciação clara de como cada tipo de SVM pode classificar um mesmo conjunto de elementos de forma diversa, dependendo da configuração de seus hiperparâmetros. Na Figura 10, é adotada uma SVM linear para a classificação entre dois conjuntos de elementos e na Figura 11, esse mesmo conjunto é classificado por uma SVM com  $Kernel RBF$  ([ALBON, 2017](#)). Para esse problema a SVM do tipo não linear mostrou-se a mais adequada para a minimização do erro de classificação.

## 3 Materiais e Métodos

O absenteísmo é um fenômeno que possui diversas características, como as citadas no Capítulo 1, em que segundo [Gosselin, Lemyre e Corneil \(2013\)](#), uma das principais teorias para explicar o absenteísmo do trabalho, seria o paradigma da escolha da decisão em que o indivíduo consciente ou inconscientemente opta por fazer-se ou não presente no ambiente laboral. Logo, buscamos considerar essas características usando dados que possam indicar a incidência de fatores que levariam ou não o agente de segurança pública a se afastar de suas atividades.

Para tanto, foram coletados dados de profissionais da Polícia Militar de Alagoas, que é um típico exemplo de instituição de segurança pública brasileira e, como tal, possui elevados índices de afastamento por adoecimento de seus agentes. Como exemplo podemos citar que apenas no ano de 2017 foram afastados de suas atividades, por motivo de adoecimento, um total de 2.055 agentes com uma média de 28,38 dias de afastamento. Esse número de agentes afastados representam 31,99% do efetivo da instituição.

Esses dados foram obtidos do sistema denominado SisHosp<sup>1</sup> que gerencia esse tipo de afastamento. Essa instituição possui, além desse sistema unificado de registros de afastamentos de agentes por adoecimento, um sistema de Recursos Humanos que gerencia as informações pessoais e funcionais dos agentes. Assim, foi solicitada a autorização junto à Polícia Militar de Alagoas para que fossem utilizados esses dados para que subsidiassem a pesquisa e que os resultados dessa pudessem retornar para a melhoria da instituição.

### 3.1 Pré-processamento dos dados

Na instituição escolhida, foi realizada a extração dos dados relacionados a afastamentos por motivo de adoecimento do sistema denominado SisHosp. Para a realização desse estudo foram utilizados os registros de afastamentos referentes ao período de 6 anos, compreendidos entre 2012 e 2017.

Também fizemos a extração dos dados do sistema denominado RH<sup>2</sup> também da Polícia Militar de Alagoas. Desse sistema foram extraídas características pessoais dos agentes e institucionais. Essas poderiam influenciar na opção do agente em praticar, consciente ou inconscientemente, o absenteísmo de longa duração. Esses dados foram adicionados a um banco de dados relacional<sup>3</sup> para que realizássemos as etapas de integração dos dados e seleção dos atributos que seriam utilizados no estudo, conforme Figura 12.

<sup>1</sup> Sistema de Gestão de Hospital da PMAL

<sup>2</sup> Sistema de Recursos Humanos da PMAL

<sup>3</sup> SGBD Postgres 9.4

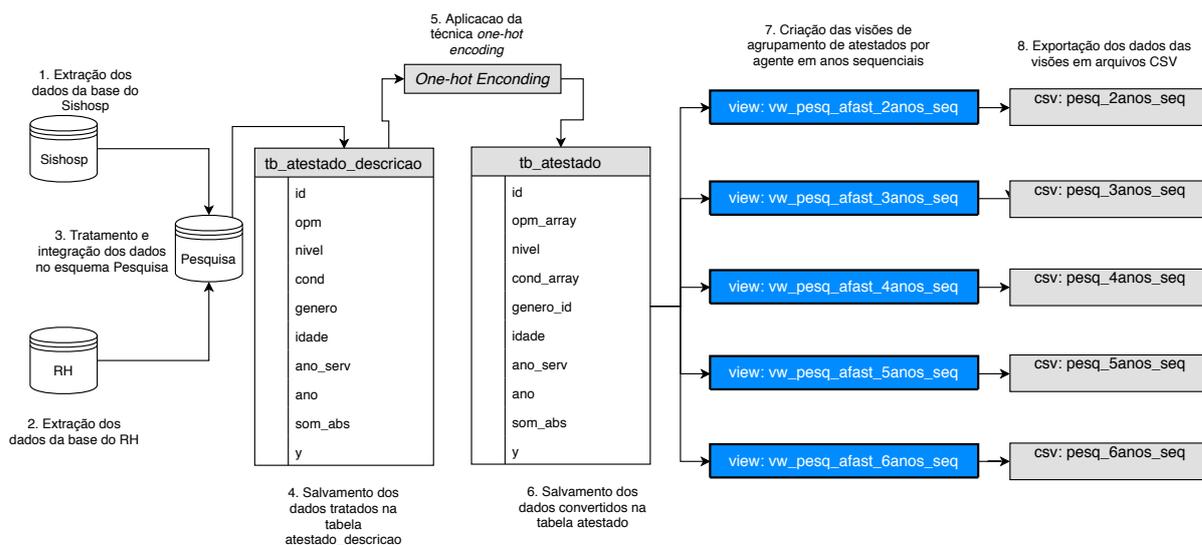


Figura 12 – Processo de extração dos dados dos sistemas de Recursos Humanos e de Gestão Hospitalar da PMAL

Fonte: autor (2018)

Assim, buscamos selecionar características relevantes para prever os agentes que têm alta probabilidade de se afastar dos que não têm, num mesmo período de tempo. Foram então selecionados como atributos para o estudo:

- O local em que o agente de segurança pública estava trabalhando quando foi afastado (OPM);
- O nível hierárquico dele na instituição (NÍVEL);
- A condição dele na atividade, se está ativo, agregado (à disposição de outra entidade), ou se já se encontra aguardando aposentadoria (COND);
- O gênero (GÊNERO);
- A idade no ano coletado (IDADE);
- Quantos anos de serviço possuía (ANOS\_SERV);
- O ano em que foi registrado o afastamento em referência (ANO);
- O somatório de dias de afastamento durante o ano (SOM\_ABS).

Esses dados representam a situação do agente de segurança pública em determinado ano em específico, ou seja, para cada agente, tem-se uma entrada para cada ano em estudo, de 2012 a 2017.

Após esse processamento, foi adicionado também um rótulo binário representando a situação do agente em determinado ano com relação a seus afastamentos por adoecimento,

respeitando a seguinte regra: se o somatório de seus afastamentos durante um determinado ano foi igual ou superior a 30 dias, o agente receberá o rótulo 1, caso contrário, receberá o rótulo 0 para o campo  $y$  da tabela  $tb\_atestado$ , na Figura 12.

Outrossim, a legislação que faz referência ao afastamento do agente policial militar é a Lei nº 5.346, de 26 de maio de 1992 (ALAGOAS, 1992), que em seu artigo 101 prevê a licença para tratamento de saúde própria com duração de 30 dias, podendo essa ser renovada por iguais períodos. É importante destacar esse prazo, pois foi adotado para delimitar as duas classes de afastamentos em análise, haja vista que o acúmulo de períodos de afastamentos por motivo de adoecimento que totalizem ou ultrapassem trinta dias anuais implicam submissão do agente de segurança pública policial militar a uma junta médica de saúde para avaliar suas condições gerais de saúde e essa avaliação pode determinar o afastamento total de suas atividades até seu reestabelecimento.

O afastamento total de suas atividades representa aos agentes prejuízos de ordem financeira, posto que não podem trabalhar em escalas extras remuneradas e implicam perda de funções gratificadas por exercício de cargo superior ao seu nível hierárquico. Essa situação leva a alguns agentes, mesmo possuindo doenças que determinariam o seu afastamento, a continuar a exercer suas atividades, o que, não raros os casos, acaba tornando mais grave a situação do agente, que finalmente se afasta por períodos muito superiores ao necessário inicialmente.

Após a extração dos dados e sua integração para associar os registros de afastamentos com os dados oriundos do sistema de RH, referentes aos agentes que se afastaram e dos que não se afastaram, chegamos a um total 36.863 tuplas. Essas tuplas foram submetidas à técnica *one hot encoding* a fim de converter atributos categóricos em numéricos. Essa técnica foi aplicada aos atributos representativos do local onde o agente exercia sua atividade no momento do afastamento (OPM), à condição do agente (COND) e ao gênero (GÊNERO). Para as demais características, foi preservado o valor obtido na operação de extração e integração dos dados, visto que já eram numéricos.

### 3.1.1 Agregação dos conjuntos de treinamento

O conjunto obtido na seção anterior foi utilizado para gerar vários conjuntos de treinamento distintos, onde cada conjunto contém instâncias representando, para cada agente, dados concatenados de um determinado intervalo de anos, iniciando com dois anos em sequência até seis anos em sequência, sendo o último ano utilizado para classificar o agente como afastado ou não de acordo com a soma de dias de afastamento naquele último ano. Como exemplo, para um determinado agente, foram selecionadas as características dele no ano de 2012 e classificado entre 0 e 1 de acordo com o somatório de dias de afastamento do agente no ano de 2013, representado pela coluna  $Y$ , como mostramos na Tabela 1. Esse processo foi repetido até que fossem alinhados para cada agente cinco anos em sequência, com o classificador sendo obtido para o sexto ano.

Tabela 1 – Exemplo de dados de agentes agregados para 2 anos em sequência - 2012-2017

Dados para o ano de referência									Dados do ano de previsão		
ID	OPM	NÍVEL	COND	GÊNE-RO	IDADE	ANO SERV	ANO	SOM ABS	ANO	SOM ABS	Y
226	DP	Maj	Ativo	1	40	19	2012	0	2013	0	0
228	CORR	Cap	Ativo	1	41	18	2012	0	2013	0	0
233	3 <sup>o</sup> BPM	Cb	Ativo	1	43	18	2012	0	2013	0	0
234	BPGD	2 <sup>o</sup> Sgt	Ativo	1	49	25	2012	0	2013	0	0
238	7 <sup>o</sup> BPM	Cb	Ativo	1	41	21	2012	0	2013	0	0
241	6 <sup>o</sup> BPM	3 <sup>o</sup> Sgt	Ativo	1	45	23	2012	0	2013	0	0
242	7 <sup>o</sup> BPM	Cb	Ativo	1	43	21	2012	0	2013	0	0
745	1 <sup>o</sup> BPM	Cb	Ativo	1	47	25	2012	150	2013	365	1

Fonte: autor (2018)

Tabela 2 – Frequência de conjuntos de treinamento por ano em sequência rotulados como absentes ou não absentes - 2012-2017

Anos em sequência	No. de absentes	No. de não absentes	Total	% de absentes
2	3.528	27.909	31.437	11,22
3	2.985	22.512	25.497	11,71
4	2.412	17.438	19.850	12,15
5	1.690	11.961	13.651	12,38
6	826	5.712	6.538	12,63

Fonte: autor (2018)

Após a agregação dos conjuntos de treinamento chegamos às quantidades de instâncias por conjunto descritas na Tabela 2, onde também está a relação percentual entre a quantidade de instâncias representativas de absenteísmo de longa duração por adoecimento e o total de tuplas por ano em sequência. Notamos que, como existiam mais possibilidades para menos anos em sequência, os conjuntos da parte superior da Tabela 2 são maiores que os da região inferior.

Os conjuntos dispostos na Tabela 2 foram usados posteriormente para avaliar se o uso de dados acumulados por um período mais longo resultaria em um preditor melhor. Foi considerado para a rotulação apenas o último ano da observação, logo, um agente que pode ter obtido 31 dias de afastamento num ano intermediário da amostra, se no último ano se afastou por apenas 5 dias, será rotulado como 0 (zero).

Os dados presentes na Tabela 2 já revelam a proporção existente entre os agentes que se afastaram e os que não se afastaram por longos períodos. Segundo os dados coletados, observamos que há um acréscimo nessa proporção à medida que mais anos acumulados de

características são considerados para cada agente.

Com base nesses dados foi possível realizar o balanceamento da amostra para que essa possuísse o mesmo quantitativo de tuplas de cada uma das duas classes. Passamos então a submeter esses dados tratados a diversas arquiteturas de aprendizagem profunda a fim de experimentar a viabilidade e robustez de diversos classificadores binários.

### 3.1.2 Configuração da variabilidade de hiperparâmetros para as arquiteturas selecionadas

Com a finalidade de avaliar quais seriam as melhores arquiteturas de redes neurais profundas para a solução do problema, fizemos uma extensa experimentação de hiperparâmetros para cada uma das arquiteturas consideradas. Para cada agregação de anos em sequência foram feitas todas as variações de hiperparâmetros descritas nas próximas seções, e para cada variação obtida, realizamos 10 (dez) repetições do experimento para se aferir os melhores resultados de cada configuração, o que resultou em 30.000 (trinta mil) experimentos, conforme descrito na Tabela 3.

Para a arquitetura do tipo Máquina de Vetores de Suporte foram também feitas as variações de hiperparâmetros, conforme descrito na Seção 3.1.2.4, sendo 160 (cento e sessenta) variações, que foram repetidas para os 5 (cinco) agrupamentos de anos em sequência e cada variante foi executada 10 (dez) vezes, como nas arquiteturas de aprendizagem profunda. Como resultado, foram 8.000 (oito mil) experimentos para essa arquitetura.

Alguns hiperparâmetros foram comuns a todas as arquiteturas, como o *batch size* que foi definido para 64 (sessenta e quatro). A função de discretização dos resultados da classificação realizada pelas arquiteturas é a descrita no Algoritmo 1, a quantidade de épocas de treinamento foi definida em 10 (dez) iterações e a função de otimização foi o Adam (KINGMA; BA, 2014).

Tabela 3 – Quantitativo de experimentos por arquitetura e de variações de hiperparâmetros - 2018

No. Camadas	MLP	RNN	LSTM
1	0	3	3
2	9	9	9
3	27	27	27
4	81	81	81
5	243	0	0
Total de Variações	360	120	120
Anos em sequência (5x)	1.800	600	600
Repetições (10x)	18.000	6.000	6.000
<b>TOTAL</b>			<b>30.000</b>

Fonte: autor (2018)

**Algorithm 1** Algoritmo para discretização

---

**Input:**  $y$   
**Output:**  $y'$   
**Result:** Retorna o rótulo binário da classe  
**if**  $y \geq 0.5$  **then**  
  | *return*1  
**else**  
  | *return*0  
**end**

---

Para cada iteração do experimento foi gerada uma matriz de confusão para avaliar os resultados obtidos por cada variante de arquitetura. Nesse sentido, os classificados corretamente como absentes pelas arquiteturas foram denominados *Verdadeiros Positivos (VP)*, os erroneamente classificados como absentes foram denominados *Falsos Positivos (FP)*. Os resultados corretamente classificados como não absentes foram denominados *Verdadeiros Negativos (VN)*, os erroneamente classificados como não absentes foram denominados *Falsos Negativos (FN)*.

Desses valores retornados, calculamos alguns indicadores para serem utilizados na comparação das arquiteturas, como a *Sensibilidade* ou *Taxa de Positivos Verdadeiros (TPV)* e a *Especificidade* ou *Taxa de Negativos Verdadeiros (TNV)*. A primeira é definida como a proporção entre os exemplos positivos corretamente previstos pelo modelo e o total de positivos, calculada conforme Equação 3.1. A segunda é a proporção entre os exemplos negativos corretamente previstos pelo modelo e o total de negativos, calculada conforme Equação 3.2 (PANG-NING; STEINBACH; KUMAR, 2009).

$$TPV = \frac{VP}{(VP + FP)} \quad (3.1)$$

$$TNV = \frac{VN}{(VN + FN)} \quad (3.2)$$

Para o cálculo da acurácia da arquitetura (*AC*), fizemos a média entre os valores obtidos pela *Sensibilidade* e *Especificidade* de cada variante, conforme Equação 3.3.

$$AC = \frac{TPV + TNV}{2} \quad (3.3)$$

Para a submissão dos dados agregados para cada configuração de hiperparâmetros, realizamos a validação cruzada desses do seguinte modo: foi selecionado randomicamente de cada grupo o percentual de 70% para que fossem utilizados no treinamento das arquiteturas e os demais 30% restantes foram utilizados para a validação. Para a variação dos hiperparâmetros da arquitetura, optamos pela técnica denominada de busca em grid, que consiste na busca exaustiva dentre todas as combinações de hiperparâmetros, selecionando aquelas que resultaram nos melhores valores para a métrica adotada para avaliação da arquitetura. No nosso caso, adotamos

a *Acurácia (AC)* como descrita na Equação 3.3 como definidora da melhor configuração de hiperparâmetros para os experimentos. Assim, descrevemos as variações de hiperparâmetros a que foram submetidas as arquiteturas avaliadas neste trabalho nas seções seguintes.

### 3.1.2.1 *Multilayer Perceptron*

Para as arquiteturas do tipo *Multilayer Perceptron* (MLP) alguns ajustes nos dados foram realizados, como a concatenação de dados anuais consecutivos, como descrito na Subseção 3.1.1 e a eliminação de colunas que representavam valores repetidos como o *ID*, *ANO* e *GÊNERO* e para o último ano só foi usado o rótulo final *y* para a classificação do exemplo. O quantitativo de neurônios avaliado por camada foi 128, 256 e 512. Além disso, o experimento também fez a variação do número de camadas, partindo de 2 (duas) a até 5 (cinco) camadas, cobrindo todas as combinações possíveis com esses valores, seguidas de uma camada de saída com 1 (um) neurônio.

A função de ativação da camada de entrada e das camadas ocultas foi definida como a *Rectified Linear Unit (ReLU)*. A função de perda adotada foi a *binary crossentropy*. Como otimizador foi adotado o Adam (KINGMA; BA, 2014). Para a camada de saída da arquitetura foi estabelecido que essa teria apenas um neurônio que classificaria entre os valores 0 (zero) e 1 (um) visto que estamos tratando de um problema de classificação binária. Para tanto, como função de ativação dessa camada foi definida a função sigmoide, e os dados de saída foram discretizados pelo Algoritmo 1.

### 3.1.2.2 *Recurrent Neural Network*

Para avaliar a arquitetura do tipo *Recurrent Neural Network* (RNN), os dados de cada ano foram arranjados de forma sequencial e por agente, para que os dados de cada ano fossem as entradas sequenciais da rede, e a classificação final *y* fosse dada após os dados do último ano serem processados pela rede.

A variação de seus hiperparâmetros consistiu na mudança do número de camadas que variou entre 1 (uma) a 4 (quatro). O número de células em cada camada variou nos valores 128, 256 e 512, o otimizador utilizado foi o Adam (KINGMA; BA, 2014), a função de perda foi a *Mean Absolute Error* (MAE). A função de discretização dos resultados da rede foi a descrita no Algoritmo 1. Para a camada de saída, representada por uma camada densa formada por um neurônio, como na MLP, foi selecionada a função de ativação sigmoide.

### 3.1.2.3 *Long Short-term Memory*

A metodologia de variação das arquiteturas do tipo *Long Short-term Memory* (LSTM) foi semelhante àquela adotada para as RNN. Os principais hiperparâmetros selecionados foram o quantitativo de camadas que também variou entre 1 (uma) até 4 (quatro) camadas, o quantitativo de células avaliadas em cada camada foi de 128, 256 ou 512.

O otimizador utilizado foi o Adam (KINGMA; BA, 2014), a função de perda utilizada foi a *Mean Absolute Error* (MAE). A função de discretização foi a descrita no Algoritmo 1. Para a camada de saída, representada por uma camada densa, composta por um neurônio, como na MLP, foi selecionada a função de ativação sigmoide.

#### 3.1.2.4 Máquina de Vetores de Suporte

Para realizar a comparação dos resultados obtidos pelas arquiteturas de aprendizagem profunda com uma técnica de aprendizagem de máquina clássica, adotamos a técnica denominada Máquina de Vetores de Suporte (SVM) com *kernel* não linear, que são aquelas capazes de promover a separação de conjuntos de dados que não estariam distribuídos linearmente. O *Kernel* usado em nossos experimentos foi *Radial Basis Function* (RBF), como na Equação 2.23.

Para a calibração dos hiperparâmetros do *Kernel* optamos pela variação do hiperparâmetro denominado  $\gamma$ , variando entre 0.1 e 2.0, com incremento em 0.1 a cada iteração, já para o hiperparâmetro de custo  $C$  adotamos os valores 0.001, 0.01, 0.1, 1, 10, 100, 1000, 10000 para cada variante.

Os dados de entrada foram concatenados da mesma forma que para as arquiteturas do tipo MLP. Para o último ano da sequência, apenas a coluna  $y$  foi utilizada, servindo de rótulo da amostra.

## 3.2 Detalhes de implementação

Para a pesquisa utilizamos a linguagem de programação Python 3.6<sup>4</sup>, a suíte de aplicativos para *Data Science* Anaconda<sup>5</sup>, o pacote para construção de redes neurais de alto nível Keras<sup>6</sup> tendo como *backend* a biblioteca TensorFlow<sup>7</sup>(ABADI; AL, 2015), o sistema gerenciador de banco de dados relacional PostgreSQL 9.4<sup>8</sup> e o software de interface gráfica para PostgreSQL pgAdmin3<sup>9</sup>. Também foram selecionadas outras bibliotecas específicas para o tratamento dos dados e geração de gráficos como Numpy<sup>10</sup>, Pandas<sup>11</sup>, Matplotlib<sup>12</sup>, Scikit-learning<sup>13</sup>, Jupyter Notebook<sup>14</sup>. O sistema operacional utilizado foi o GNU/Linux Ubuntu 18.04 em um computador Intel Core i5 com 32 Gbytes de memória RAM.

<sup>4</sup> Disponível em: <<https://www.python.org/>>. Acessado em 25 ago. 2018.

<sup>5</sup> Disponível em: <<https://www.anaconda.com/>>. Acessado em 25 ago. 2018.

<sup>6</sup> Disponível em: <<https://keras.io/>>. Acessado em 26 ago. 2018.

<sup>7</sup> TensorFlow, o logotipo da TensorFlow e quaisquer marcas relacionadas são marcas registradas da Google Inc. Disponível em: <<https://www.tensorflow.org/>>. Acessado em 26 ago. 2018.

<sup>8</sup> Disponível em: <<https://www.postgresql.org/>>. Acessado em 10 out. 2018.

<sup>9</sup> Disponível em: <<https://www.pgadmin.org/>>. Acessado em 10 out. 2018.

<sup>10</sup> Disponível em: <<http://www.numpy.org/>>. Acessado em 25 ago. 2018.

<sup>11</sup> Disponível em: <<https://pandas.pydata.org/>>. Acessado em 25 ago. 2018.

<sup>12</sup> Disponível em: <<https://matplotlib.org/>>. Acessado em 25 ago. 2018.

<sup>13</sup> Disponível em: <<https://scikit-learn.org/stable/>>. Acessado em 25 ago. 2018.

<sup>14</sup> Disponível em: <<https://jupyter.org/>>. Acessado em 25 ago. 2018.

## 4 Resultados e Discussão

### 4.1 Resultados

Após uma extensa experimentação das arquiteturas selecionadas, seguindo a metodologia descrita no Capítulo 3, percebemos que a solução proposta possui viabilidade técnica para aplicação no cotidiano das instituições. Assim, após a submissão dos dados dos agentes de segurança pública às diversas arquiteturas com suas respectivas variações de hiperparâmetros, foi possível avaliar quais seriam as mais adequadas para a construção de um preditor de absentismo, como observamos a seguir.

#### 4.1.1 *Multilayer Perceptron*

As arquiteturas de redes neurais do tipo *Multilayer Perceptron*, foram experimentadas em todas as variações de hiperparâmetros como o descrito na Seção 3.1.2.1, dessas variações obtivemos os resultados dispostos na Tabela 4 referentes aos melhores índices de acurácia.

Quando analisamos os resultados referentes a 6 anos em sequência de informações de agentes (Tabela 4), obtivemos resultados superiores a 78%. Considerando a complexidade do problema em estudo, trata-se de um valor expressivo, que já permitiria uma seleção efetiva de agentes que deveriam ser encaminhados aos serviços de acompanhamento à saúde das instituições, para prevenção. Em seguida, selecionamos as dez melhores configurações encontradas, considerando todos os experimentos realizados para as redes do tipo MLP, conforme Tabela 5.

Notamos que para as dez melhores arquiteturas de redes do tipo MLP, com a variação de seus hiperparâmetros, os resultados de acurácia mais expressivos são obtidos quando as redes são alimentadas por dados de 6 (seis) anos em sequência de cada agente, conforme Tabela 5. Contudo, observamos também, que para alcançar esses resultados, a configuração dos hiperparâmetros da rede é muito mais complexa do que a encontrada em outras arquiteturas testadas (Tabelas 6, 8, 10), sendo os melhores resultados obtidos com 5 ou 6 camadas, muitas delas utilizando-se de

Tabela 4 – Melhores arquiteturas para as redes do tipo MLP por anos em sequência - 2012-2017

Anos em Sequência	No. de Camadas	Neurônios por Camada	No. de parâmetros	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
2	6	256-512-512-256-128-1	586.753	76,5977	66,0075	71,3026
3	6	128-512-128-256-512-1	324.737	77,5330	67,5737	72,5534
4	6	256-128-256-256-256-1	281.217	80,1090	65,3147	72,7118
5	6	256-512-128-128-512-1	391.425	79,0385	70,6114	74,8250
6	5	256-128-128-128-1	204.801	77,6923	79,1667	78,4295

Fonte: autor (2018)

Tabela 5 – Melhores índices de acurácia para arquiteturas do tipo MLP - 2012-2017

Anos em Sequência	No. de Camadas	Neurônios por Camada	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
6	5	256-128-128-128-1	77,6923	79,1667	78,4295
6	5	128-128-128-512-1	79,6154	77,0833	78,3494
6	6	256-128-128-128-256-1	86,8313	69,1057	77,9685
6	6	256-512-256-512-256-1	76,9231	78,2979	77,6105
6	6	256-512-128-512-128-1	86,4198	68,6992	77,5595
6	6	256-128-256-128-512-1	85,1852	69,9187	77,5519
6	6	512-512-128-256-128-1	82,7160	72,3577	77,5369
6	5	128-512-256-128-1	81,5261	73,5294	77,5278
6	5	256-512-128-128-1	86,1538	68,7500	77,4519
6	5	256-256-128-128-1	77,6923	77,0833	77,3878

Fonte: autor (2018)

Tabela 6 – Configurações das melhores arquiteturas para as redes do tipo RNN por anos em sequência - 2012-2017

Anos em Sequência	No. de Camadas	Neurônios por Camada	No. de parâmetros	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
2	3	128-512-1	359.169	82,4545	58,1888	70,3216
3	3	512-512-1	843.777	75,0556	67,1981	72,5534
4	5	256-256-128-128-1	307.329	78,1081	66,3817	72,2449
5	2	128-1	30.593	71,3461	74,8490	73,0976
6	5	128-128-256-128-1	211.329	75,2066	79,0322	77,1194

Fonte: autor (2018)

512 neurônios em cada uma, o que aumentou significativamente o número de parâmetros da rede.

#### 4.1.2 Recurrent Neural Network

Para as arquiteturas do tipo *Recurrent Neural Network* (RNN), também foram encontrados resultados satisfatórios, contudo, inferiores àqueles obtidos pelas MLP (Tabela 4). Para os melhores resultados, considerando todos os intervalos de anos em sequência, a acurácia das arquiteturas para 6 anos em sequência foi superior a 77%, como observamos na Tabela 6, o que também é bastante significativo para o problema enfrentado.

Dos dez melhores resultados obtidos pela variação dos hiperparâmetros dessa arquitetura, demonstrados na Tabela 7, podemos observar que resultaram da alimentação da rede com 6 anos de dados em sequência, e que houve grande alternância das configurações de hiperparâmetros, em que redes relativamente mais simples conseguiram índices elevados de acurácia. Acreditamos que a recursividade inerente a esse tipo de arquitetura contribuiu para a redução do número de parâmetros necessários à obtenção dos melhores índices de acurácia.

Contudo, essa característica da arquitetura RNN não foi determinante para alcançar os melhores resultados, haja vista que observamos também redes com configurações de hiperparâ-

Tabela 7 – Melhores índices de acurácia para arquiteturas do tipo RNN - 2012-2017

Anos em Sequência	No. de Camadas	Neurônios por Camada	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
6	5	128-128-256-128-1	75,2066	79,0323	77,1194
6	5	128-512-128-128-1	68,1818	83,0645	75,6232
6	2	128-1	71,6535	78,6325	75,1430
6	2	256-1	75,1969	74,7863	74,9916
6	5	128-128-128-128-1	68,5950	80,2419	74,4185
6	2	256-1	79,5181	69,1667	74,3424
6	5	512-256-128-128-1	62,3967	85,8871	74,1419
6	3	128-256-1	75,9843	72,2222	74,1032
6	3	128-128-1	77,5591	70,0855	73,8223
6	5	128-128-256-512-1	68,1818	79,4355	73,8087

Fonte: autor (2018)

metros bastante complexas ocupando as primeiras classificações como melhores arquiteturas viáveis. Quando observamos os resultados obtidos para os *Verdadeiros Positivos*, que são aqueles agentes corretamente classificados como propensos ao absentismo de longa duração, algumas dessas configurações obtiveram resultados superiores a 80%, o que seriam índices de acurácia muito satisfatórios para o problema enfrentado, mas obtidos às custas de moderadas taxas de *Verdadeiros Negativos* e elevada complexidade na configuração de seus hiperparâmetros.

Porém, isso indica que há sim um padrão determinante, que pode ser encontrado naqueles indivíduos que praticam o absentismo de longa duração e que estes estão relacionados às características coletadas nos sistemas das instituições. Observamos também que não são esses fatores exclusivamente os responsáveis pelos afastamentos prolongados, mas que contribuiriam para a prática desse comportamento nas instituições.

### 4.1.3 Long Short-term Memory

As arquiteturas de redes do tipo *Long Short-term Memory* (LSTM) obtiveram resultados expressivos quando receberam como entrada dados de 6 (seis) anos em sequência. Nesses casos, obtivemos para o índice de acurácia valores superiores a 76%, conforme dados da Tabela 8. Esses valores foram conseguidos com configurações de hiperparâmetros relativamente mais simples que àquelas experimentadas em outras arquiteturas, onde os melhores resultados foram obtidos apenas com duas camadas, possuindo cada uma, respectivamente 128 células e 1 neurônio para classificação. Com isso, o número de parâmetros necessários à configuração da rede também foi muito diminuto.

Para os dez melhores resultados obtidos pela variação dos hiperparâmetros da arquitetura, conforme Tabela 9, observamos que essa tendência permaneceu onde a arquitetura formada por apenas duas camadas, com 128 células na primeira camada e um neurônio para classificação, mostrou-se bastante eficaz na solução do problema.

Tabela 8 – Configurações das melhores arquiteturas para as redes do tipo LSTM por anos em sequência - 2012-2017

Anos em Sequência	No. de Camadas	Células por Camada	No. de parâmetros	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
2	4	512-128-128-1	1.733.761	84,0344	54,9504	69,4924
3	4	128-128-512-1	1.566.721	70,4439	72,5298	71,4868
4	5	128-128-512-128-1	1.894.529	71,7421	70,8100	71,2760
5	5	128-128-128-128-1	516.737	73,2931	70,7129	72,0030
6	2	128-1	121.985	74,3295	78,4090	76,3692

Fonte: autor (2018)

Tabela 9 – Melhores índices de acurácia para arquiteturas do tipo LSTM - 2012-2017

Anos em Sequência	No. de Camadas	Células por Camada	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
6	2	128-1	74,3295	78,4091	76,3693
6	2	128-1	75,8621	76,1364	75,9992
6	2	256-1	77,3946	73,8636	75,6291
6	2	128-1	76,2452	75,0000	75,6226
6	3	128-128-1	74,3295	76,8939	75,6117
6	3	256-128-1	73,5632	77,2727	75,4180
6	2	128-1	72,0307	78,7879	75,4093
6	4	512-256-256-1	77,7344	72,8745	75,3044
6	3	128-128-1	74,7126	75,7576	75,2351
6	2	128-1	73,1801	77,2727	75,2264

Fonte: autor (2018)

Destacamos que, novamente, os melhores índices de acurácia foram obtidos para as configurações em que a rede é alimentada por dados de 6 (seis) anos em sequência. Contudo, a média dos resultados para a acurácia foi de 75%, mesmo para dados acumulados de mais anos (Tabela 9).

Novamente, o que se destaca nessa arquitetura, que utiliza células densamente conectadas e de forma recursiva, é a redução significante da complexidade da rede para se obter valores semelhantes às demais arquiteturas. Mesmo não obtendo os melhores índices, essa arquitetura demonstrou ser promissora para esse tipo de problema, haja vista sua capacidade de armazenar informações para relacioná-los aos dados anteriormente recebidos e aos dados que ainda serão carregados. Assim, os resultados obtidos por essa arquitetura ratificam que o acúmulo de dados de mais anos em sequência melhorariam a acurácia do preditor.

Logo, quando observados isoladamente os resultados das redes obtidos por quantitativo de anos em sequência, há um incremento considerável na precisão da rede, como se vê na Tabela 8, em que para 2 (dois) anos em sequência a precisão é de 69,49%, já quando analisamos o melhor resultado para uma rede alimentada com dados de agentes referentes a 6 (seis) anos em sequência, essa precisão aumenta para 76,36%.

Tabela 10 – Configurações das melhores arquiteturas para as SVM por anos em sequência - 2012-2017

Anos em Sequência	Kernel	$\gamma$	$C$	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
2	rbf	0,1	1	76,5977	66,0075	71,3026
3	rbf	0,1	1	69,4503	71,6352	70,5427
4	rbf	0,1	1	60,3283	80,8333	70,5808
5	rbf	0,1	1	52,1401	86,3179	69,2290
6	rbf	0,1	1	49,6063	90,6383	70,1223

Fonte: autor (2018)

Tabela 11 – Melhores índices de acurácia para arquiteturas do tipo SVM - 2012-2017

Anos em Sequência	Kernel	$\gamma$	$C$	Verdadeiro Negativo (%)	Verdadeiro Positivo (%)	Acurácia (%)
4	rbf	0.1	1	60,3283	80,8333	70,5808
3	rbf	0.1	1	69,4503	71,6352	70,5427
4	rbf	0.4	1	59,3707	81,2500	70,3104
3	rbf	0.1	1	70,1856	70,2890	70,2373
6	rbf	0.1	1	49,6063	90,6383	70,1223
3	rbf	0.1	1	69,1704	70,7118	69,9411
6	rbf	0.1	1	48,7903	90,7489	69,7696
2	rbf	0.1	1	80,0939	59,0328	69,5634
2	rbf	0.1	1	78,7645	60,3189	69,5417
6	rbf	0.1	10	48,4252	90,6383	69,5317

Fonte: autor (2018)

#### 4.1.4 Máquina de Vetores de Suporte

Os modelos de aprendizado de máquina do tipo Máquinas de Vetores de Suporte (SVM) conseguiram resultados satisfatórios para as melhores configurações de hiperparâmetros experimentadas. Com suas taxas de acurácia atingindo 70% quando receberam como entrada os dados referentes a 6 (seis) anos em sequência, conforme observamos na Tabela 10. Destacamos que em relação às arquiteturas de redes neurais profundas, foi a que obteve os menores índices de acurácia.

Outro ponto que destacamos é que as configurações de hiperparâmetros presentes nos melhores resultados, mesmo com a alternância de intervalos de anos, praticamente não sofreram alterações. Para o *Kernel* do tipo *RBF*, os valores do hiperparâmetro  $\gamma$  foram constantes em 0,1 (zero vírgula um). Para o hiperparâmetro  $C$  o melhor valor encontrado foi 1 (um), com apenas um resultado divergente dessa configuração, como observamos na Tabela 11.

Após essas observações, promovemos uma comparação mais detalhada entre os resultados obtidos por cada arquitetura e a quantidade de anos em sequência que foram recebidos como entrada e a variação de seus hiperparâmetros, como segue.

Tabela 12 – Comparação da acurácia entre as melhores arquiteturas de MLP, LSTM, RNN e SVM por anos em sequência - 2012-2017

Anos em Sequência	MLP (%)	LSTM (%)	RNN (%)	SVM (%)
2	71,3026	69,492	70,3216	69,5634
3	72,5534	71,487	71,1269	70,5427
4	72,7118	71,276	72,2449	70,5808
5	74,8250	72,003	73,0976	69,2290
6	78,4295	76,369	77,1194	70,1223

Fonte: autor (2018)

#### 4.1.5 Comparação entre os resultados obtidos pelas arquiteturas selecionadas

Seguindo com os experimentos, fizemos uma comparação direta entre os melhores resultados obtidos por cada arquitetura. Nessa comparação, vista na Tabela 12, destacamos as redes neurais artificiais do tipo *Multilayer Perceptron* que conseguiram os melhores resultados em relação às demais arquiteturas, mesmo quando consideramos as variações de anos em sequência e de hiperparâmetros.

Na Tabela 12 estão dispostos os melhores resultados em relação à acurácia, obtidos por cada arquitetura e para cada grupo de anos em sequência.

Fizemos também uma análise de como os resultados obtidos por cada combinação de hiperparâmetro de cada arquitetura estariam distribuídos para cada grupo de anos acumulados. Utilizamos para isso o diagrama de caixa por permitir visualizar essa variação nos dados de forma mais objetiva, pois uma arquitetura poderia apresentar os melhores resultados mas esses obtidos por valores atípicos ou *outliers*.

Para tanto, na Figura 13 mostramos uma comparação direta entre os índices de acurácia obtidos por cada arquitetura experimentada considerando a quantidade de 2 (dois) e 3 (três) anos acumulados. Percebemos que para a arquitetura do tipo MLP, apesar de possuir uma quantidade expressiva de *outliers*, possui uma concentração de resultados em intervalos em valores superiores aos das demais arquiteturas e mais condensados, mesmo em comparação aos acumulados de dois e de três anos.

A arquitetura do tipo SVM apresentou um intervalo bem amplo em relação as demais arquiteturas, sobretudo quando observados os resultados obtidos para 3 (três) anos acumulados (Figura 13b). Para 3 (três) anos acumulados, como observamos na Figura 13b praticamente há alguma variação de SVM em todo o intervalo aceito nos experimentos.

Quando observamos as medidas de acurácia obtidas para 4 (quatro) e 5 (cinco) anos acumulados, percebemos que a acurácia obtida pelas arquiteturas do tipo MLP ainda permanecem superiores as demais arquiteturas, contudo, as arquiteturas do tipo LSTM e RNN tiveram uma melhora significativa em seus resultados como está evidente na análise da Figura 14ab.

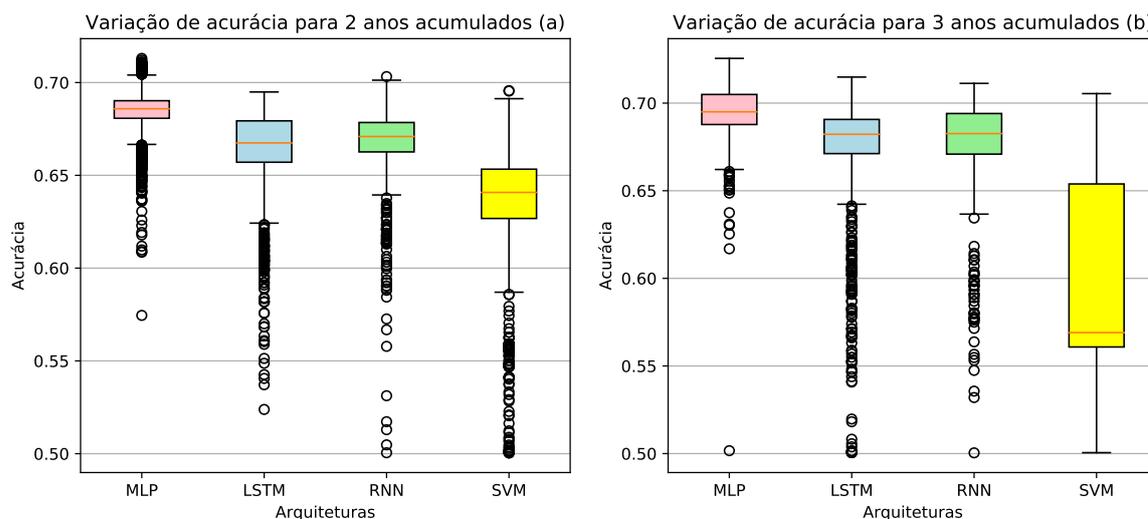


Figura 13 – Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 2 anos acumulados (a); para 3 anos acumulados (b) - 2012-2017

Fonte: autor (2018)

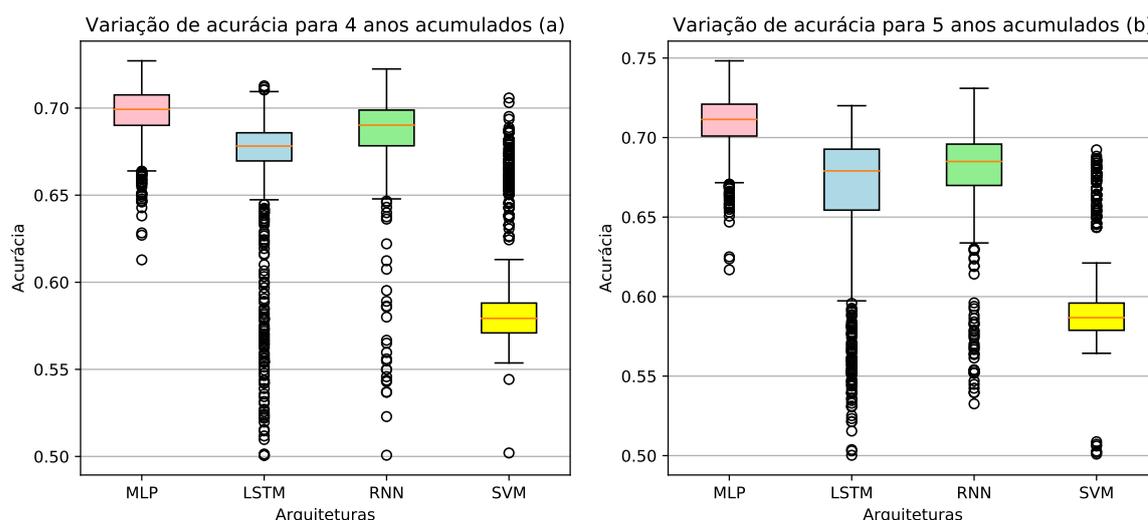


Figura 14 – Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 4 anos acumulados (a); para 5 anos acumulados (b) - 2012-2017

Fonte: autor (2018)

Entretanto, percebemos uma piora nos valores de acurácia obtidos pelas arquiteturas do tipo SVM que além de muitos *outliers*, seus principais valores se concentraram em intervalos mais baixos, como se observa na Figura 14ab. Isso pode ser consequência da maior dimensão dos dados.

Continuando a análise, quando as arquiteturas são alimentadas com dados de 6 (seis) anos acumulados, as variações de cada arquitetura sofreram um decréscimo em seus intervalos de acurácia, contudo, algumas dessas variações, mesmo que *outliers*, alcançaram valores significativos, resultando em acurácias superiores aos 75%. Essa situação foi comum as três

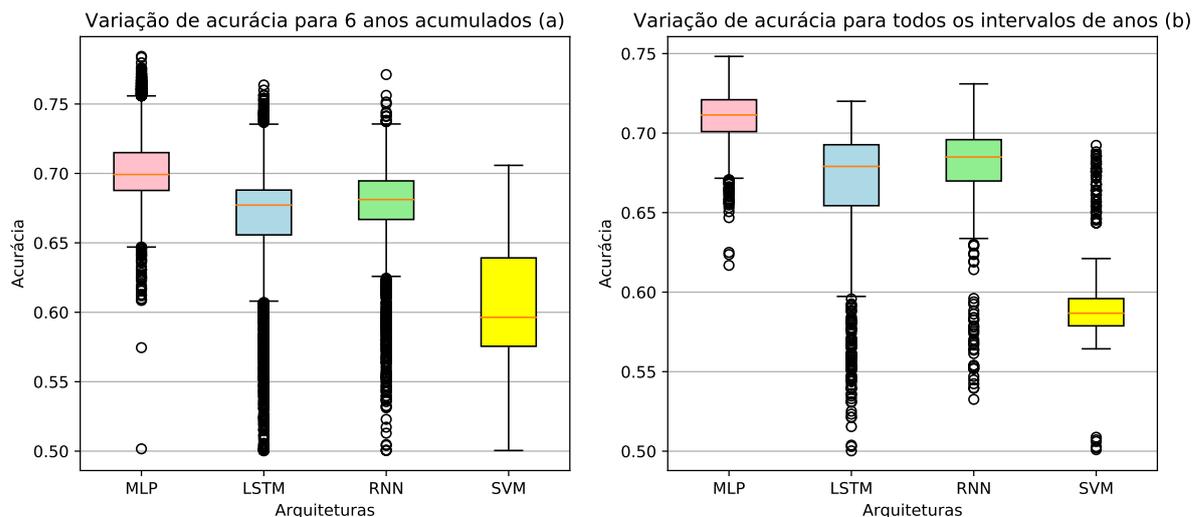


Figura 15 – Gráfico comparativo das medidas de dispersão das acurácias considerando a quantidade de anos acumulados: para 6 anos acumulados (a); para todos os intervalos de anos acumulados (b) - 2012-2017

Fonte: autor (2018)

arquitecturas de aprendizado profundo, como está evidente na Figura 15a.

Para visualizarmos os resultados obtidos de uma forma geral, também fizemos a comparação entre todos os resultados obtidos pelas arquiteturas para todos os intervalos de anos acumulados considerados, o que pode ser visto na Figura 15b. Nela podemos notar que mesmo considerando todas as variações de hiperparâmetros e a variação de dados de anos acumulados, as arquiteturas do tipo MLP ainda possuem os melhores resultados quando analisadas por esse tipo de abordagem.

Em seguida fizemos uma comparação entre as melhores acurácias obtidas por cada arquitetura e a quantidade de anos acumulados, para avaliarmos qual seria a tendência desses resultados, o que está expresso na Figura 16.

Com isso, observamos que existe uma tendência de aumento da acurácia das arquiteturas quando essas são alimentadas com dados de mais anos acumulados em sequência de cada agente. Essa tendência só não é observada do mesmo modo quando analisamos os resultados obtidos para a arquitetura do tipo SVM. Nela, para 5 (cinco) anos de dados de anos em sequência, houve um decréscimo no índice de acurácia. Em particular, houve significativo aumento da acurácia das redes RNN e LSTM quando dados de 6 anos foram usados. Isso pode ser um indício de que, com o acúmulo de dados de mais anos, estas arquiteturas possam superar as MLP.

Contudo, ficou demonstrada a influência do quantitativo maior de dados de anos em sequência em quase todas as arquiteturas para a melhoria dos resultados (Figura 16). Notamos também que os índices de acurácia para as arquiteturas experimentadas, estão, com exceção da SVM, em valores superiores a 70%, o que comprova a relação entre os dados utilizados na pesquisa e sua relação com os elevados índices de absentismo experimentados pelos agentes de

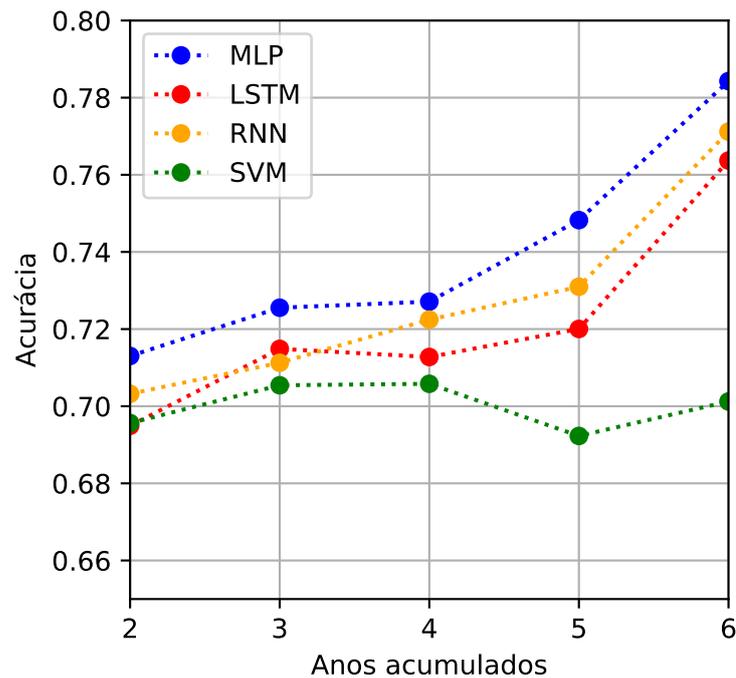


Figura 16 – Gráfico da acurácia das melhores arquiteturas, considerando a quantidade de anos acumulados: eixo horizontal representa a quantidade de anos e o eixo vertical a melhor acurácia - 2012-2017

Fonte: autor (2018)

segurança pública.

Nesse sentido, quando comparados os resultados de todas as variações de arquitetura e de hiperparâmetros obtidos nos experimentos, observamos que algumas arquiteturas seriam mais adequadas à predição do absenteísmo de longa duração do que outras. Isso porque, para algumas arquiteturas, grande parte de suas variantes, obtiveram índices muito abaixo do esperado, então resolvemos construir histogramas para que fosse possível visualizar melhor a relação entre cada configuração e a quantidade de variantes por intervalo de acurácia.

Para a comparação, fizemos uma relação percentual entre a quantidade de variantes de cada arquitetura e a acurácia obtida por essa variante. Na Figura 17, para dados de 2 (dois) anos em sequência, vê-se que as redes do tipo MLP possuem mais variantes que atingiram índices de acurácia próximas ou superiores a 70%, enquanto às demais ficaram com grande parte de suas variações abaixo desse índice, chegando inclusive a valores aproximados de 50%, o que praticamente indica que não conseguiram classificar entre as instâncias.

Quando realizada essa comparação para as variantes das arquiteturas alimentadas com dados de 3 (três) anos em sequência, observamos uma aproximação dos percentuais, percebidas principalmente para as arquiteturas do tipo LSTM e RNN. Entretanto, os índices de acurácia obtidos pelas variantes da arquitetura do tipo MLP ainda continuam concentrados em valores próximos a 70%, conforme Figura 18.

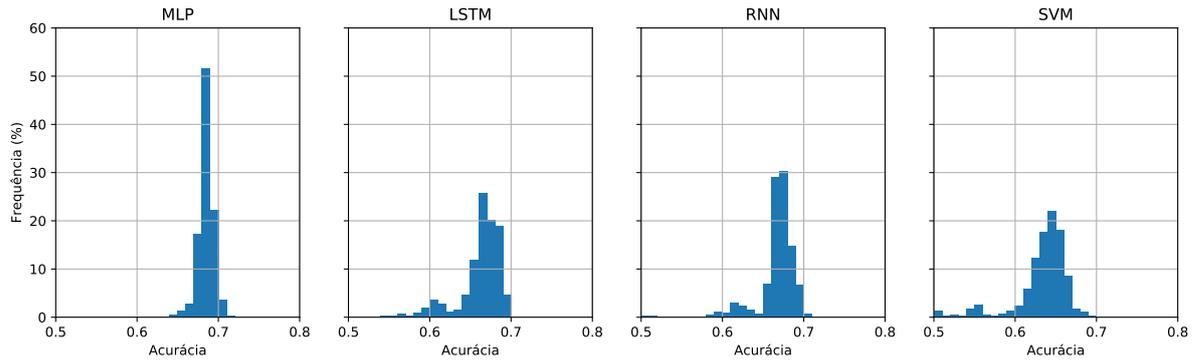


Figura 17 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 2 anos em sequência - 2012-2017  
 Fonte: autor (2018)

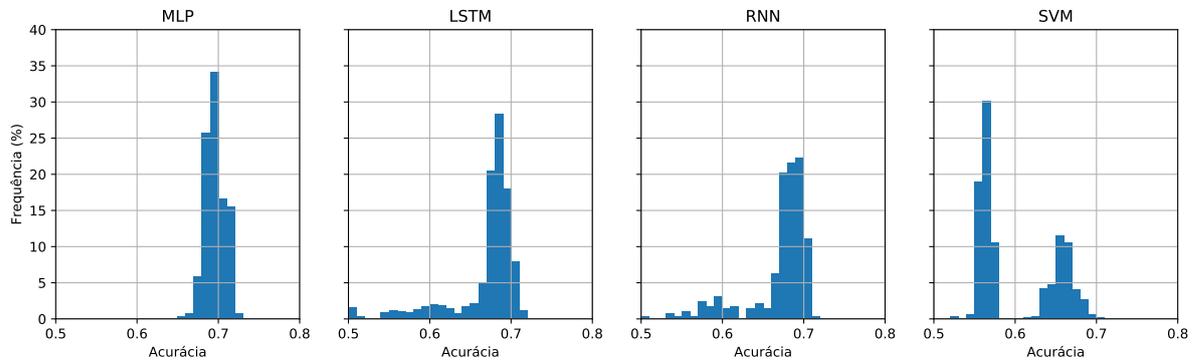


Figura 18 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 3 anos em sequência - 2012-2017  
 Fonte: autor (2018)

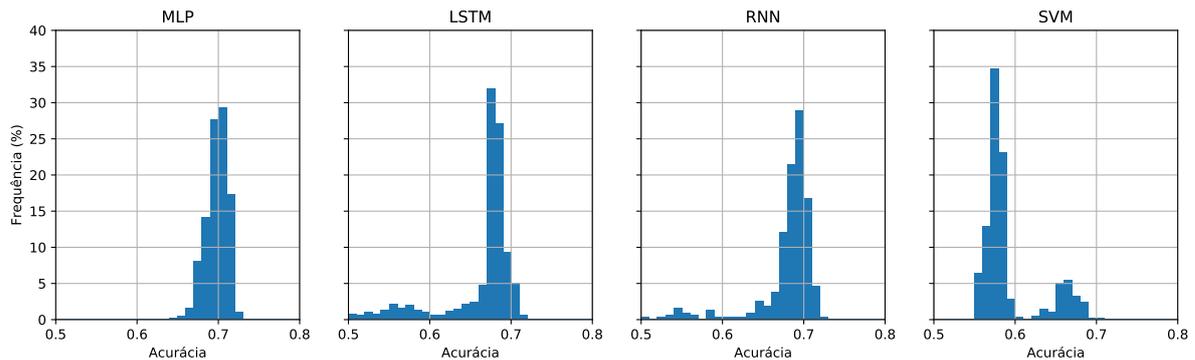


Figura 19 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 4 anos em sequência - 2012-2017  
 Fonte: autor (2018)

Seguimos o experimento e comparamos a acurácia obtida pelas variantes das arquiteturas agora alimentadas com dados de 4 anos em sequência. Nessa situação, as variantes conseguiram melhoras significativas nos seus índices de acurácia, onde para as redes do tipo MLP, percebemos uma concentração de resultados distribuídos agora sobre os 70%. Para as redes do tipo LSTM, apesar de algumas variantes apresentarem resultados inferiores aos 60%, houve um acréscimo

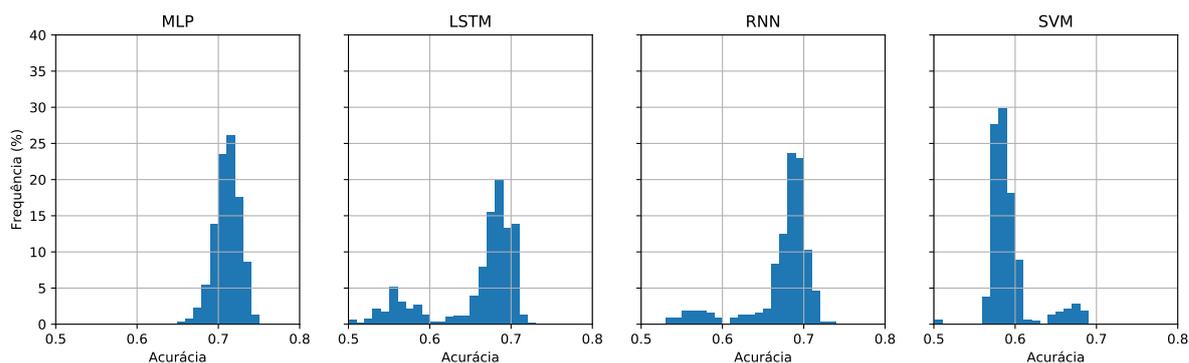


Figura 20 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 5 anos em sequência - 2012-2017

Fonte: autor (2018)

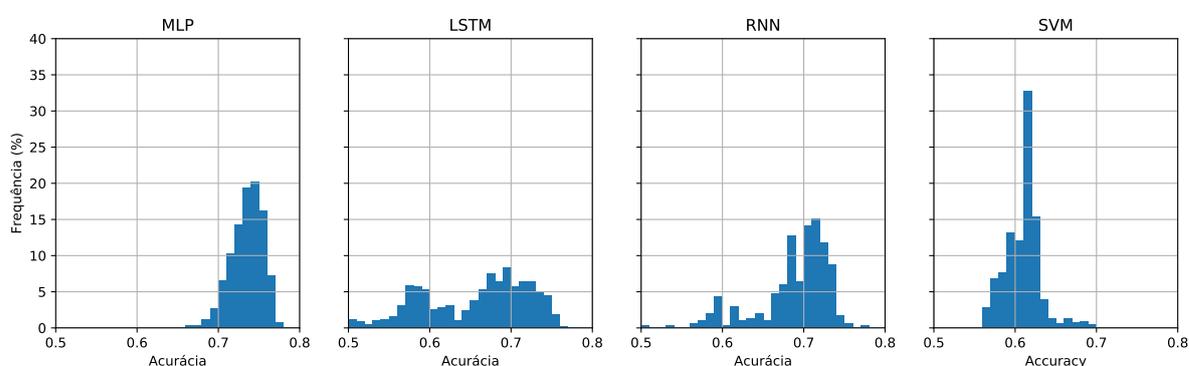


Figura 21 – Histogramas das acurácias obtidas pelas variantes de cada arquitetura quando recebem como entrada dados de 6 anos em sequência - 2012-2017

Fonte: autor (2018)

expressivo de variantes que se aproximaram dos 68% de acurácia. Essa situação também é observada para as variantes do tipo RNN, em que houve uma concentração de variantes com acurácias próximas aos 68%. Contudo, para as arquiteturas do tipo SVM, nessas condições, houve um decréscimo nas acurácias registradas pelo experimento, em que um percentual significativo de suas variantes obteve índices inferiores aos 58%, conforme observamos na Figura 19.

Quando realizamos a comparação das variantes para as arquiteturas que foram alimentadas com dados de 5 anos em sequência, conforme Figura 20, destacamos novamente a concentração de variantes da arquitetura do tipo MLP que obtiveram índices de acurácia superiores a 70%, bem como o decréscimo na acurácia das arquiteturas do tipo LSTM. Para as arquiteturas do tipo RNN, as variantes não foram influenciadas significativamente pela alteração do quantitativo de anos em sequência. Novamente, quando observamos os resultados obtidos pelas variantes baseadas em SVM, vimos que houve uma significativa diminuição do número de variantes que obtiveram índices elevados de acurácia, tendo uma quantidade expressiva dessas obtendo índices abaixo dos 60%, conforme Figura 20.

Para as variantes alimentadas com os dados referentes a 6 anos em sequência, novamente as arquiteturas baseadas em redes do tipo MLP destacaram-se por apresentar uma concentração

percentual expressiva de variantes com índices de acurácia superiores aos 70%. Para as arquiteturas do tipo LSTM, houve uma distribuição desse percentual por todos os níveis de acurácia, observado pela distribuição de suas incidências pela linha base da Figura 21. Para as variantes baseadas na arquitetura RNN, percebemos que houve uma melhora significativa nos seus índices, em que, grande parte das suas variantes atingiram valores superiores a 70%. Para as arquiteturas do tipo SVM também foi registrada melhora nos seus índices, contudo, permanecendo em sua maioria, com valores aproximados a 63%.

Isso demonstra que os resultados obtidos por cada arquitetura foram significativamente afetados pela variação dos hiperparâmetros e pela quantidade maior de anos em sequência para os dados de entrada de cada experimento. Dependendo dessa entrada, algumas configurações passaram a obter índices de acurácia superiores a 70%. Para as arquiteturas do tipo RNN observamos que, com o aumento do número de anos acumulados, um quantitativo maior de variantes obteve índices de acurácia superiores a 70%. Já para as variantes de arquiteturas baseadas em SVM, observamos um movimento oposto, no qual praticamente todas essas obtiveram valores inferiores aos 70% de acurácia, inclusive o maior percentual dessas variantes concentrou-se na região compreendida entre os 55% e 64%, conforme é possível notarmos na Figura 21.

#### 4.1.6 Visualização dos dados pela técnica de projeção multidimensional *T-Distributed Stochastic Neighbor Embedding*

Para visualizar a relação existente entre as características coletadas e o absenteísmo de longa duração dos agentes de segurança pública, adotamos como subsídio os dados referentes a 6 anos em sequência, justamente por terem conseguido os melhores resultados independente da arquitetura selecionada.

Como ferramenta de projeção utilizamos a técnica *T-Distributed Stochastic Neighbor Embedding* (t-SNE) de autoria de [Maaten e Hinton \(2008\)](#). Essa técnica de projeção multidimensional permite a visualização de dados em alta dimensão por meio da criação de projeções de cada dado em uma representação bidimensional ou tridimensional desses dados.

Essa técnica tem como vantagem sobre outras formas de representações a facilidade de otimização e a produção de melhores visualizações dos dados pela redução da tendência de agrupamento de pontos no centro do mapa criado por outras técnicas ([MAATEN; HINTON, 2008](#)).

Com isso, aplicamos o t-SNE e obtivemos a projeção 2D mostrada na Figura 22, onde estão marcadas em vermelho a projeção das características dos indivíduos propensos ao absenteísmo de longa duração, e em azul, aqueles que não estariam propensos a esse tipo de afastamento. Nessa projeção, notemos que são destacadas como regiões de características suscetíveis, aquelas localizadas na margem inferior esquerda da projeção e como não suscetíveis as da margem superior esquerda (Figura 22). Logo, indivíduos que possuíssem suas características

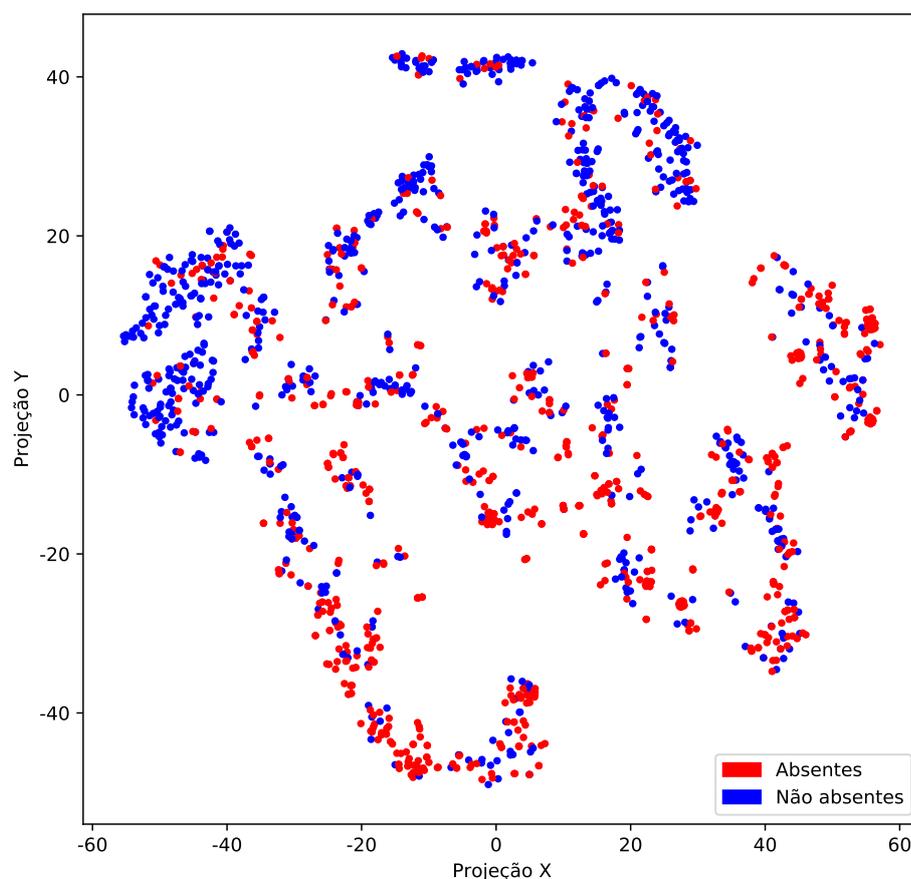


Figura 22 – Projecção t-SNE 2D para absentes e não absentes para dados de 6 anos em sequência - 2012-2017

Fonte: autor (2018)

voltadas para a porção inferior da projeção, seriam aqueles que estariam mais sujeitos à prática do absenteísmo de longa duração motivado por adoecimento e que deveriam ser encaminhados aos setores de acompanhamento de saúde das instituições.

Contudo, fica mais claro com essa projeção, a complexidade do problema em estudo, haja vista existirem regiões onde essas características estão praticamente sobrepostas, como na região central da Figura 22. Isso mostra que os padrões de absenteísmo não são triviais, e que nosso melhor preditor, que atingiu índices superiores a 78% de acurácia, é excelente quando considerada a complexidade do problema e de suas variáveis.

Logo, passamos à exploração da visualização dos dados pela projeção em três dimensões para avaliar se seria possível observar regiões espaciais distintas entre as características presentes nos dados de agentes submetidos às arquiteturas. Avaliamos se o padrão de separação também ocorreria numa projeção tridimensional, para tanto, continuamos marcando as projeções referentes aos agentes que se afastam por períodos iguais ou superiores a trinta dias com pontos vermelhos e em azul aquelas projeções referentes aos dados de agentes que se afastam por períodos inferiores a 30 dias, conforme se observa na Figura 23.

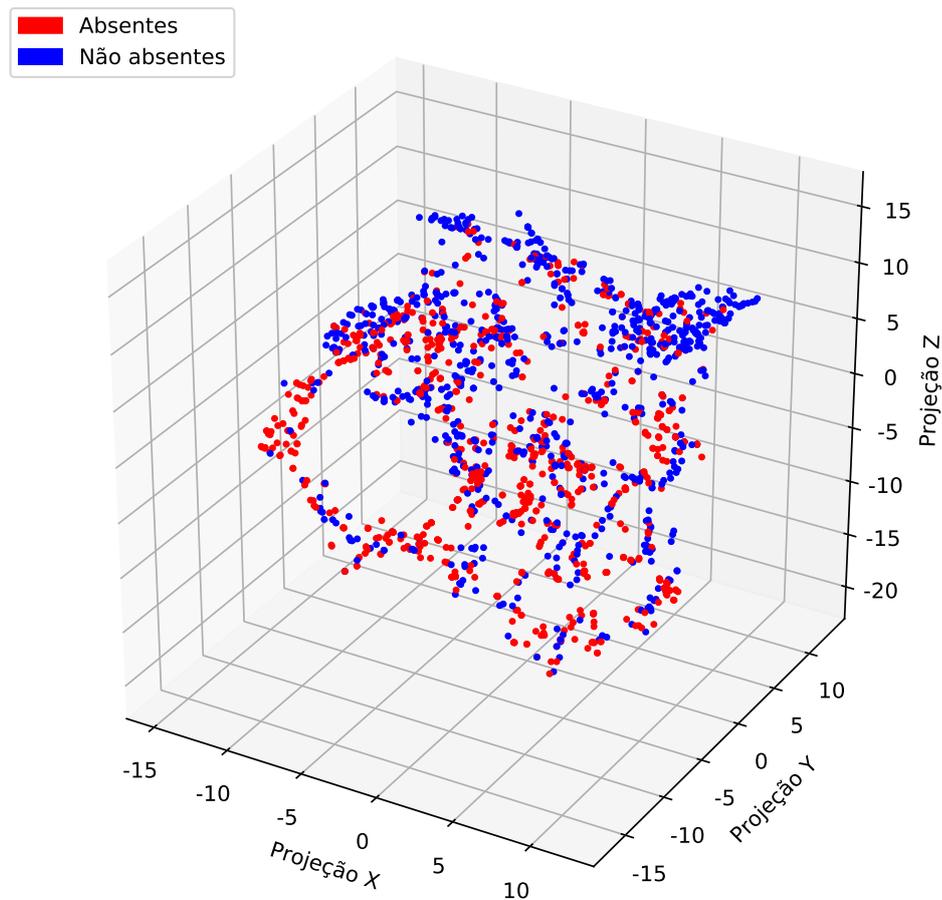


Figura 23 – Projeção t-SNE 3D para absentes e não absentes para 6 anos em sequência - 2012-2017

Fonte: autor (2018)

Como visto, novamente os dados apresentam elevada complexidade observada por sua representação tridimensional pela técnica t-SNE. Fizemos então a simplificação dessas características pela aplicação da técnica de redução de dimensionalidade denominada *Principal Component Analysis* (PCA), que realiza uma projeção para um determinado espaço de dimensão  $n$  (PHUONG; LIN; ALTMAN, 2005). No nosso caso, foi estabelecido  $n = 50$  a fim de se obter uma simplificação da exibição dos dados e a consequente construção da projeção bidimensional usando novamente a técnica t-SNE. Podemos observar o resultado na Figura 24.

Assim, após a aplicação da técnica de visualização multidimensional t-SNE (MAATEN; HINTON, 2008) na projeção por PCA (PHUONG; LIN; ALTMAN, 2005), observamos a complexidade do problema em análise, que, sob o ponto de vista dos dados submetidos aos experimentos, não há uma distinção clara das características intrínsecas que estariam relacionadas ao absentéismo de longa duração dos agentes de segurança pública, haja vista não haver, na projeção utilizada, a delimitação de regiões em que essa situação estaria mais evidente, como observamos na Figura 24. Ao contrário, existem predominantemente, regiões em que as características estão praticamente sobrepostas. Desse modo, passamos então ao estudo da relação

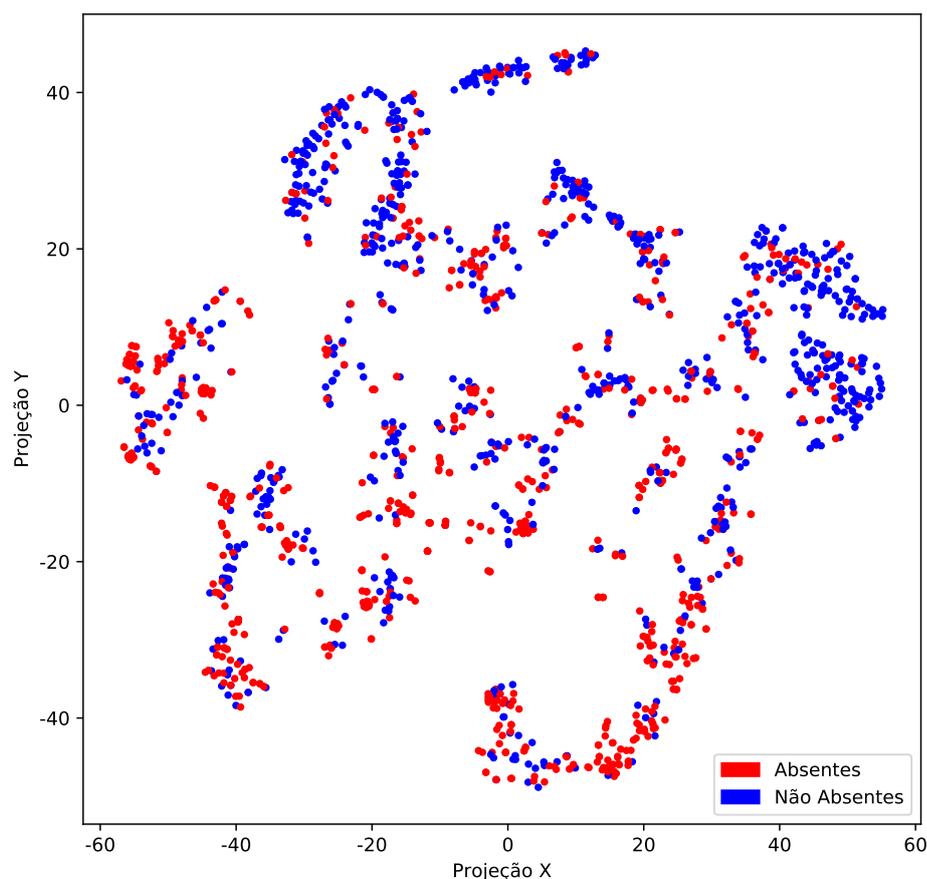


Figura 24 – Projecção t-SNE 2D para absentes e não absentes para 6 anos em sequência com redução de dimensionalidade por PCA - 2012-2017

Fonte: autor (2018)

entre essas características e o absenteísmo por um método *Wrapper*.

#### 4.1.7 Análise da influência e relação entre atributos por um método *Wrapper*

A partir dos resultados obtidos pela visão multidimensional dos dados, buscou-se analisar quais dos atributos ou conjunto de atributos utilizados no estudo influenciariam na determinação das características que levariam os agentes de segurança pública a se afastarem de suas atividades por longos períodos, para tanto, os atributos envolvidos na análise dos dados foram submetidos ao método de seleção de atributos do tipo *Wrapper*, que realiza a seleção exaustiva dentre todos os subconjuntos de um grupo de atributos selecionados (PHUONG; LIN; ALTMAN, 2005).

Esses subconjuntos de atributos foram utilizados na alimentação da arquitetura que conseguiu os melhores índices de acurácia nos experimentos, qual seja, a configuração da arquitetura do tipo MLP, com 5 (cinco) camadas densamente conectadas e alimentadas com dados de 6 (seis) anos em sequência e com o quantitativo de neurônios por camada de 256 – 128 – 128 – 128 – 1, respectivamente (Tabela 5).

Tabela 13 – Resultado de acurácia para a seleção de atributos por um método *Wrapper* para MLP de 5 camadas e tendo como entrada os dados de 6 anos em sequência - 2012-2017

No. de características	Subgrupos de características	Verdadeiro negativo (%)	Verdadeiro positivo (%)	Acurácia (%)
5	('OPM', 'COND', 'IDADE', 'ANO_SERV', 'SOM_ABS')	82,9787	74,8971	78,9379
4	('OPM', 'COND', 'GÊNERO', 'SOM_ABS')	76,1702	79,0123	77,5913
6	('OPM', 'COND', 'GÊNERO', 'IDADE', 'ANO_SERV', 'SOM_ABS')	86,0169	68,4015	77,2092
5	('COND', 'GÊNERO', 'IDADE', 'ANO_SERV', 'SOM_ABS')	86,4407	67,6580	77,0493
6	('OPM', 'NÍVEL', 'COND', 'GÊNERO', 'ANO_SERV', 'SOM_ABS')	84,2553	69,5473	76,9013
6	('OPM', 'NÍVEL', 'COND', 'IDADE', 'ANO_SERV', 'SOM_ABS')	86,4407	66,9145	76,6776
5	('COND', 'GÊNERO', 'IDADE', 'ANO_SERV', 'SOM_ABS')	78,8235	74,3902	76,6069
4	('COND', 'GÊNERO', 'IDADE', 'SOM_ABS')	79,6078	73,5772	76,5925
4	('COND', 'IDADE', 'ANO_SERV', 'SOM_ABS')	88,1356	64,3123	76,2239
4	('OPM', 'COND', 'ANO_SERV', 'SOM_ABS')	79,1489	73,2510	76,2000
7	Todas as características	87,7119	64,6840	76,1979

Fonte: autor (2018)

Logo, na Tabela 13, descrevemos os 11 (onze) melhores resultados obtidos pela aplicação do método para a seleção de atributos que foram submetidos à melhor arquitetura e configuração retornadas nos experimentos. Nesse resultado, observamos que alguns atributos se destacam e estão presentes nas variações que obtiveram os melhores resultados, como os atributos que indicam o local onde o agente exerce suas atividades (OPM), o tempo em que ele está na instituição (ANO\_SERV) e o quantitativo de dias de afastamento a que ele se submeteu nos anos anteriores até o afastamento prolongado no último ano da amostra (SOM\_ABS). Notamos que o conjunto completo das características só aparece na Tabela 13 após 10 (dez) resultados de variações pelo método *Wrapper*.

Diante disso, percebemos a necessidade de avaliar, dentre todas as características utilizadas no experimento, quais desses atributos foram mais frequentes nos resultados apontados pelo método *Wrapper*, o que resultou na análise disposta na Tabela 14. Assim, nessa análise, algumas outras características tiveram destaque, como a situação do agente nos anos anteriores ao seu afastamento prolongado (COND). O somatório de dias de afastamento nos anos anteriores ao ano de classificação (SOM\_ABS), seguido pelo atributo relacionado à quantidade de anos de serviço do agente na instituição (ANO\_SERV). A idade do agente antes do afastamento prolongado (IDADE), o local onde exerceu suas atividades nos anos anteriores ao absentéismo prou-

Tabela 14 – Frequência de características para o método de seleção de atributos *Wrapper* para os 10 melhores resultados retornados para MLP de 5 camadas e tendo como entrada os dados de 6 anos em sequência - 2012-2017

Característica	Frequência	%
COND	10	20,41
SOM_ABS	10	20,41
ANO_SERV	8	16,33
IDADE	7	14,29
OPM	6	12,24
GÊNERO	6	12,24
NÍVEL	2	4,08
TOTAL	49	100,00

Fonte: autor (2018)

gado (OPM), o gênero do agente (GÊNERO) e por último seu nível hierárquico na instituição (NÍVEL).

Percebemos que, de fato, alguns atributos influenciam no comportamento dos agentes que se fazem presentes ou não ao ambiente laboral e que esses devem ser considerados em atividades de prevenção, como é o caso da própria situação funcional do profissional e do quantitativo de dias de afastamento a que ele vem sendo submetido por anos consecutivos, que podem servir de indicadores de alerta para possíveis incidências do fenômeno do absenteísmo de longa duração.

Com base nesses resultados, também investigamos quais unidades possuíam os índices mais elevados de afastamentos em relação ao quantitativo de agentes da instituição, considerando os 6 anos da amostra. Para esses dados, algumas unidades se destacaram por concentrarem um número expressivo de afastamentos prolongados, como observamos na Figura 25.

Na Figura 25 vimos que as unidades que possuem os maiores índices de afastamento em relação a toda a amostra utilizada no trabalho são aquelas que estão diretamente ligadas à atividade-fim da instituição pesquisada, o que pode denotar que deveriam ser esses os principais alvos de campanhas de prevenção ao absenteísmo de longa duração. Contudo, continuamos o estudo e fizemos uma comparação entre o quantitativo existente de agentes em cada unidade e seus índices de afastamentos prolongados registrados, o que podemos observar na Figura 26.

Observamos que houve uma clara mudança de posição na classificação das unidades pesquisadas, surgindo unidades administrativas, que *a priori* não seriam consideradas numa eventual campanha de prevenção ao absenteísmo de longa duração se fossem considerados apenas valores absolutos. Na Figura 26, a unidade denominada EMG possui nos anos utilizados para os experimentos, mais de 50% de seu efetivo afastado por longos períodos, seguido também da unidade denominada SEDS, formada pelos agentes que prestam serviço na Secretaria de Segurança Pública do Estado de Alagoas. Em terceiro lugar, encontramos a unidade denominada PM4, responsável pela gestão orçamentária da instituição pesquisada. Só em quarto lugar na comparação relativa é que aparece a primeira unidade responsável pela atividade-fim da

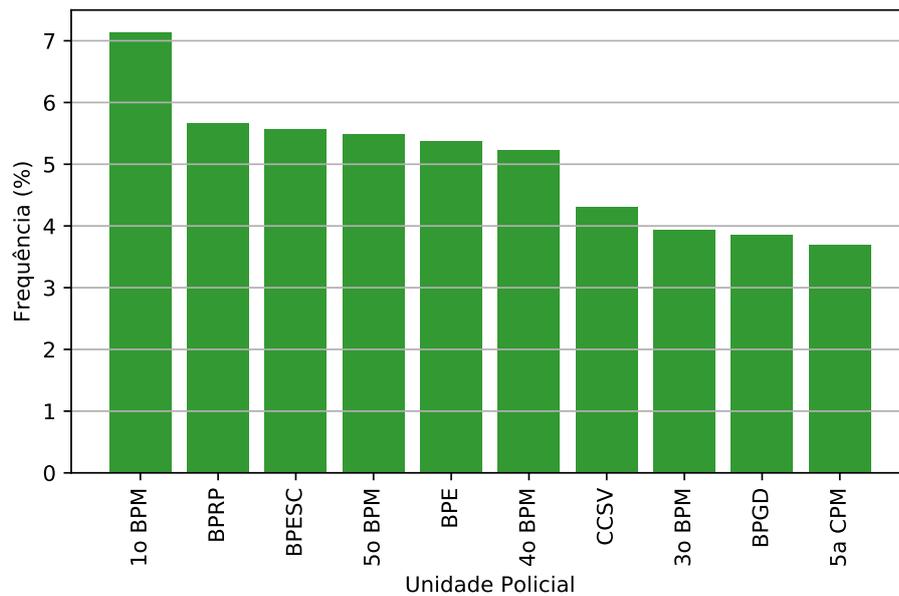


Figura 25 – Frequência percentual relativa de absenteísmo de longa duração considerando o número de afastamentos superiores a 30 dias de toda a instituição - 2012-2017  
 Fonte: autor (2018)

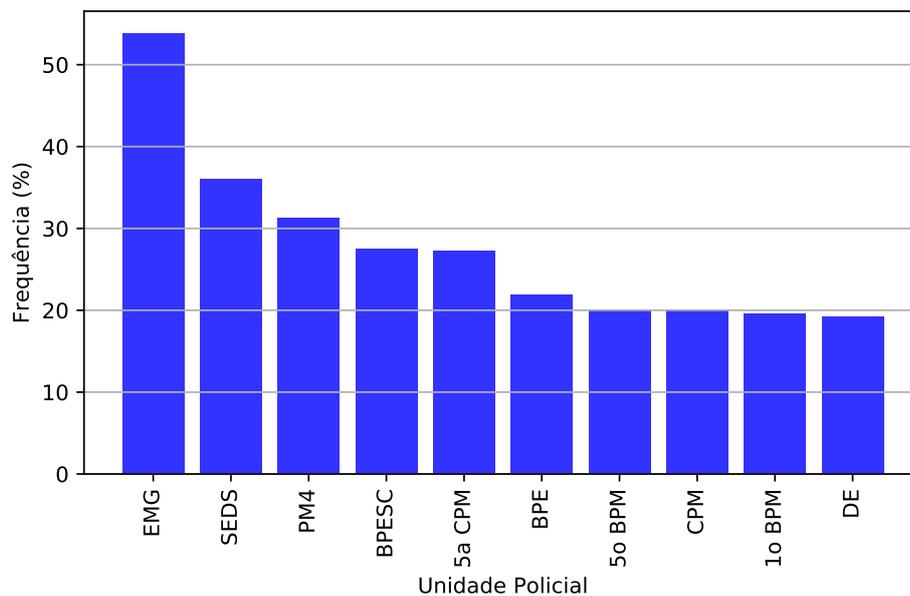


Figura 26 – Frequência percentual relativa de absenteísmo de longa duração considerando o número de afastamentos superiores a 30 dias em relação ao efetivo de cada unidade - 2012-2017  
 Fonte: autor (2018)

instituição, denominada BPESC.

Logo, o fenômeno do absenteísmo motivado por adoecimento, mais uma vez, revela-se muito complexo e que uma análise superficial dos dados poderia indicar uma direção totalmente adversa ao enfrentamento eficaz da problemática, ressaltando a necessidade da adoção de técnicas

que realmente possam extrair o conhecimento de dados que estão sendo subutilizados pelos integrantes das instituições, muitas vezes por desconhecer os benefícios que técnicas como a aprendizagem profunda possam trazer às suas atividades cotidianas. Seguimos então para uma comparação entre os resultados obtidos por meio de técnicas que permitissem avaliar a relação entre a arquitetura e os resultados obtidos, optando pela Curva Característica de Operação dos Receptores (ROC).

#### 4.1.8 Análise comparativa das arquiteturas por Curvas Características de Operação dos Receptores

Com os resultados obtidos pela aplicação do método *Wrapper* sobre a melhor configuração de arquitetura obtida nos experimentos, resolvemos promover uma análise comparativa entre os resultados obtidos para às demais arquiteturas sob as mesmas circunstâncias. Para fazer esse comparativo, utilizamos a técnica conhecida como Curva de Características de Operação de um Receptor (ROC).

Uma curva ROC é uma metodologia para visualização do balanceamento entre as taxas obtidas por um determinado classificador, realizando uma comparação entre os resultados classificados como *Positivos Verdadeiros* e *Positivos Falsos*. Nessa representação, a taxa de positivos verdadeiros é projetada no eixo  $y$  do gráfico, enquanto a taxa de *Positivos Falsos* é projetada no eixo  $x$ . Cada ponto que compõe a curva ROC representa os resultados obtidos pelo classificador (PANG-NING; STEINBACH; KUMAR, 2009).

Outro dado que pode ser extraído da visualização dos resultados obtidos por um classificador por meio de sua projeção por meio de uma curva ROC é a área sob a curva (AUC), que indicaria, *a priori*, qual modelo classificaria melhor em média. Pelo modelo, uma arquitetura que realizasse a classificação perfeita apresentaria uma  $AUC = 1$ , já um modelo que classificasse de forma aleatória teria uma  $AUC = 0,5$ . Nesse sentido, uma maior área sob a curva (AUC) representaria o melhor modelo segundo o classificador adotado (PANG-NING; STEINBACH; KUMAR, 2009).

Partimos desse princípio para projetar uma curva ROC para os resultados obtidos por cada arquitetura. Sendo selecionadas aquelas que receberam como entrada os dados de 6 (seis) anos em sequência e nas configurações que obtiveram os melhores resultados de acurácia. Além disso, também dispomos na projeção os resultados produzidos pela arquitetura do tipo MLP para a seleção de atributos realizada pelo método *Wrapper* que obteve os melhores resultados, conforme Seção 4.1.7. O resultado foi a projeção das curvas ROC dispostas na Figura 27.

Após essa projeção (Figura 27), observamos que as curvas ROC para as arquiteturas do tipo MLP, LSTM, RNN e MLP com a aplicação de um método de seleção de atributos *Wrapper* são muito semelhantes, e que os classificadores selecionados estariam fazendo um trabalho bem eficiente. Entretanto, ao continuarmos a análise, observamos uma diferenciação considerável em

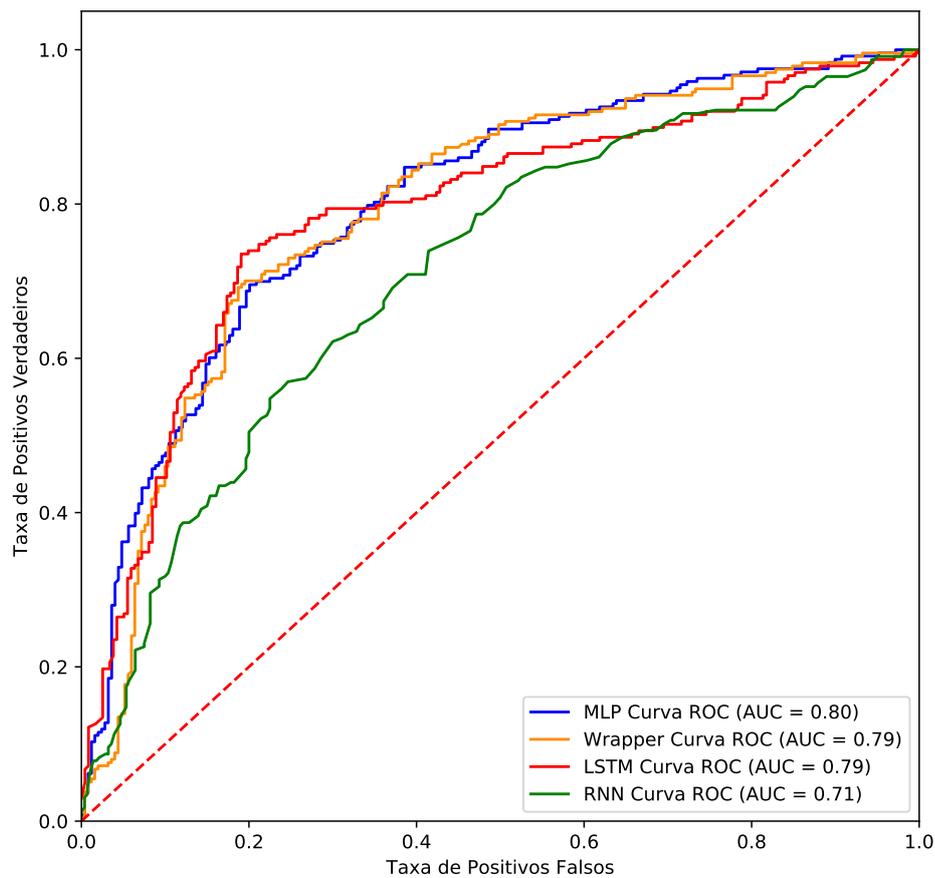


Figura 27 – Comparação entre curva ROC e AUC das melhores configurações de arquiteturas e seleção de atributos pelo método *Wrapper* para 6 anos em sequência - 2012-2017  
Fonte: autor (2018)

relação ao outro parâmetro nesse tipo de projeção, que é a área sob a curva (AUC).

Considerando que esse parâmetro representa os agentes que foram corretamente classificados como propensos à prática de absenteísmo de longa duração. Observamos claramente que a arquitetura do tipo MLP estaria sendo mais eficiente em classificar esse tipo de situação, haja vista que conseguiu uma AUC de 80%, enquanto sua variante obtida pela seleção de atributos por um método *Wrapper* obteve uma AUC de 79%. Esse mesmo valor foi obtido pela arquitetura do tipo LSTM que registrou 79% de AUC. Já a arquitetura do tipo RNN obteve uma AUC de 71% (Figura 27).

Novamente, por meio dessa representação, ficou comprovado que para o problema proposto as redes neurais do tipo *Multlayer Perceptron* foram as que mais conseguiram se aproximar de um classificador realmente eficaz, conseguindo inferir com índices aproximados a 80% os agentes que estariam ou não predispostos à prática de absenteísmo de longa duração, com índices dessa significância, poderiam ser estabelecidos períodos de análise dos dados dos agentes para que fossem selecionados para intervenção pelos órgãos de apoio à saúde nas instituições.

## 4.2 Discussão

A problemática abordada é de complexidade elevada, haja vista o número de pesquisas relacionadas às mais diversas áreas dedicadas à interpretação, análise e construção de metodologias para o combate a esse tipo de comportamento, que causa prejuízos de toda ordem, tanto às instituições, como aos agentes públicos e por último, a toda a sociedade que perde com a não efetivação dos resultados esperados pelos agentes de segurança pública.

Outrossim, os agentes de segurança pública, relutam em procurar o acompanhamento necessário ao tratamento de diversos problemas de saúde física e psicológica que teriam ligação com o comportamento absente, permitindo que esses males se agravem, findando com o afastamento prolongado de suas atividades. Já os órgãos responsáveis pelo acompanhamento de saúde das instituições, geralmente contam com um efetivo diminuto frente ao público a que se destinam. Pesa ainda o acometimento de um número expressivo de indivíduos com a problemática, o que ainda acarreta uma sobrecarga do serviço, dificultando ainda mais as ações preventivas efetivas.

Ressaltamos que estamos tentando prever com o trabalho, a real possibilidade de um indivíduo, agente de segurança pública, dentro de todas as variáveis que o cerca, a chance de ele praticar o absenteísmo de longa duração motivado por adoecimento, seja esse adoecimento causado por problemas físicos ou psicológicos. Esse problema, por si só, é demasiadamente complexo, o que explicaria a variação nos resultados obtidos pelas arquiteturas experimentadas.

Apesar disso, como observado, podemos inferir que os resultados provaram que a utilização de técnicas de aprendizagem profunda podem sim contribuir para a prevenção do absenteísmo de longa duração, provocado principalmente, por afastamentos relacionados ao adoecimento desses agentes, com índices de acerto superiores a 78%, o que já permitiria a construção de indicadores efetivos de possíveis agentes propensos a esse tipo de comportamento. Vimos também que, a simples estatística sobre os dados de afastamentos dos agentes de segurança pública iria trazer resultados diversos daqueles que representam os reais indivíduos que necessitam de acompanhamento e também não permitiria individualizá-los com base em características que já estão guardadas nos sistemas das instituições.

Assim, se pudermos fornecer uma ferramenta efetiva e que permita aos setores de acompanhamento da saúde, identificar com antecedência os agentes de segurança pública, diante mesmo de sintomas iniciais ou características comuns àqueles que se afastam por longos períodos, isso traria além do resgate da saúde desses profissionais e sua autoestima, a correta aplicação de recursos públicos, que passariam a ser destinados ao público interno que realmente estaria sob o risco da prática do absenteísmo de longa duração. Além disso, abriríamos as portas para que as mais modernas metodologias e tecnologias adentrassem esses órgãos de segurança, não só para sua atividade-fim, mas para a manutenção do seu principal patrimônio, que são os profissionais de segurança pública.

# Conclusões e Trabalhos Futuros

## Conclusões

Nesse trabalho fizemos uma comparação exaustiva entre as arquiteturas de redes neurais dos tipos *Multilayer Perceptron*, *Recurrent Neural Network*, *Long-short Term Memory* e da técnica de aprendizagem de máquina denominada Máquina de Vetores de Suporte, com a variação da configuração de seus hiperparâmetros. Avaliamos quais seriam as mais adequadas para a construção de um preditor de absenteísmo de longa duração de agentes de segurança pública. Após o emprego dessas técnicas, concluímos que a rede neural do tipo *Multilayer Perceptron* é a mais adequada à solução do problema, pois apresentou taxas de acurácia superiores a 78%, apesar da complexidade do problema, demonstrada pela projeção dos dados pela técnica t-SNE.

Fizemos a avaliação de quais características institucionais ou individuais estariam mais relacionadas ao fenômeno do absenteísmo de longa duração, chegando à conclusão de que o local onde o agente exerce suas atividades, o histórico de afastamentos anteriores e o tempo que exerce a atividade são características comuns aos agentes que se afastam por longos períodos. A fim de permitir um melhor entendimento sobre o problema, identificamos quais unidades possuiriam, proporcionalmente, os maiores índices de absenteísmo de longa duração.

Com isso, ficou demonstrado que é possível a construção de preditores de absenteísmo motivados por adoecimento com os dados disponíveis nos sistemas das instituições pela utilização de técnicas de aprendizagem profunda. Pelas características dos dados utilizados, verificamos que as redes artificiais do tipo *Multilayer Perceptron* ainda seriam as mais indicadas para esse tipo de problema, porém não esgotando o tema que ainda carece de muitas pesquisas para avaliar se também não seria possível juntar a esse instrumento, outros citados nos demais trabalhos a respeito do tema, como questionários aplicados aos agentes de forma sazonal e avaliações da organização como um todo.

## Trabalhos Futuros

Como trabalho futuro, propomos a execução de novos experimentos, incluindo novas características intrínsecas aos agentes de segurança pública e outras características da instituição. Adicionaríamos a essa nova entrada a avaliação de um questionário para coletar dados de outros fatores que também podem influenciar no absenteísmo de longa duração, como condição familiar, quantidade de filhos, estado civil, relacionamento com pares e superiores, satisfação com o trabalho etc, tudo sob supervisão dos setores de psicologia das instituições.

Será também feita a busca por novas ferramentas de visualização dos dados a fim

de clarear o entendimento sobre esse fenômeno complexo. Essas novas ferramentas poderão demonstrar a relação existente entre os diversos fatores intrínsecos e extrínsecos ao indivíduo, motivadores do comportamento absente nas instituições de segurança pública e fomentar novas abordagens para o enfrentamento do problema.

Outra atividade será a construção de um software que quando somado as suítes de softwares gerenciais das instituições, permitiria de forma automática, a realização de todo o processo de extração dos dados, o tratamento, a integração e a geração de gráficos e indicadores como os mostrados no presente trabalho. Além disso, esse software teria como saída a lista de agentes de segurança que deveriam ser encaminhados aos setores de atenção à saúde física ou psicológica para acompanhamento, isso quando fossem detectadas as características comuns aos profissionais que praticam o absenteísmo de longa duração.

As técnicas avaliadas e a experiência obtida no presente trabalho, podem ainda ser estendidas para a prevenção de outros males, como a análise automática de dados de prontuários médicos, identificando quais seriam as principais relações entre as enfermidades profissionais e outras características dos indivíduos e dos locais onde esses exercem seus ofícios, permitindo intervenções antes mesmo da instalação da enfermidade.

Logo, os resultados obtidos no presente trabalho já demonstraram que a utilização de técnicas de aprendizagem profunda podem ajudar na identificação precoce de características de agentes propensos ao absenteísmo de longa duração, permitindo uma ação efetiva dos setores responsáveis pela prevenção desse tipo de fenômeno nas instituições e que essas técnicas ainda podem ser empregadas com sucesso em outros tipos de intervenções relacionadas à saúde desses profissionais.

# Referências

- ABADI, M.; AL et. *TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems*. 2015. Software available from tensorflow.org. Disponível em: <<https://www.tensorflow.org/>>. Citado na página 34.
- AFOEM. *Workplace attendance and absenteeism*. 1. ed. Sydney: Australasian Faculty of Occupational and Environmental Medicine, 1999. Disponível em: <<https://www.racp.edu.au/docs/default-source/faculties-library/workplace-attendance-and-absenteeism.pdf?sfvrsn=4>>. Acesso em: 22 abr 2019. Citado 2 vezes nas páginas 9 e 8.
- ALAGOAS. *Lei nº 5.346, de 26 de maio de 1992. Dispõe sobre o estatuto dos policiais militares do estado de alagoas e dá outras providências*. 1992. Lei nº 5.346, de 26 de maio de 1992. Disponível em: <<http://www.conselhodeseguranca.al.gov.br/legislacao/corpo-de-bombeiros-militar-de-alagoas/Lei5346-estatudo.pdf>>. Acesso em: 09 set. 2018. Citado na página 29.
- ALBON, C. *SVC Parameters When Using RBF Kernel*. 2017. Disponível em: <[https://chrisalbon.com/machine\\_learning/support\\_vector\\_machines/svc\\_parameters\\_using\\_rbf\\_kernel/](https://chrisalbon.com/machine_learning/support_vector_machines/svc_parameters_using_rbf_kernel/)>. Citado 2 vezes nas páginas 25 e 26.
- BRASIL. *Constituição da República Federativa do Brasil promulgada em 5 de outubro de 1988*. [S.l.]: Congresso Nacional, 1988. Citado na página 3.
- CHO, K. et al. Learning phrase representations using rnn encoder-decoder for statistical machine translation. *arXiv preprint arXiv:1406.1078*, 2014. Citado na página 22.
- CORNEIL, W. et al. Exposure to traumatic incidents and prevalence of posttraumatic stress symptomatology in urban firefighters in two countries. *Journal of occupational health psychology*, Educational Publishing Foundation, v. 4, n. 2, p. 131, 1999. Citado na página 4.
- FACELI et al. *Inteligência Artificial: Uma abordagem de Aprendizagem de Máquina*. Rio de Janeiro: Livros Técnicos e Científicos Editora Ltda, 2011. ISBN 9788521618805. Citado 2 vezes nas páginas 21 e 22.
- FAN, C.-Y. et al. Using hybrid data mining and machine learning clustering analysis to predict the turnover rate for technology professionals. *Expert Systems with Applications*, v. 39, n. 10, p. 8844 – 8851, 2012. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417412002473>>. Citado na página 6.
- GOODFELLOW, I. et al. *Deep learning*. [S.l.]: MIT press Cambridge, 2016. v. 1. Citado 10 vezes nas páginas 9, 10, 11, 15, 16, 17, 18, 19, 23 e 24.
- GOODFELLOW, I. et al. Generative adversarial nets. In: *Advances in neural information processing systems*. [S.l.: s.n.], 2014. p. 2672–2680. Citado na página 2.
- GOSELIN, E.; LEMYRE, L.; CORNEIL, W. Presenteeism and absenteeism: Differentiated understanding of related phenomena. *Journal of occupational health psychology*, Educational Publishing Foundation, v. 18, n. 1, p. 75, 2013. Disponível em: <<https://psycnet.apa.org/record/2012-34902-001>>. Citado 6 vezes nas páginas 1, 2, 5, 8, 9 e 27.

- HASSAN, H. et al. Assessment of artificial neural network for bathymetry estimation using high resolution satellite imagery in shallow lakes: Case study el burullus lake. *International Water Technology Journal*, v. 5, 12 2015. Disponível em: <[https://www.researchgate.net/figure/A-hypothetical-example-of-Multilayer-Perceptron-Network\\_fig4\\_303875065](https://www.researchgate.net/figure/A-hypothetical-example-of-Multilayer-Perceptron-Network_fig4_303875065)>. Acesso em: 19 abr. 2019. Citado na página 22.
- HAYKIN, S. *Redes Neurais - 2ed.* BOOKMAN COMPANHIA ED, 2001. ISBN 9788573077186. Disponível em: <<https://books.google.com.br/books?id=IBp0X5qfyjUC>>. Citado 7 vezes nas páginas 11, 12, 13, 14, 15, 16 e 17.
- HOMAIE-SHANDIZI, A.-H. et al. Flight deck crew reserve: From data to forecasting. *Engineering Applications of Artificial Intelligence*, v. 50, p. 106 – 114, 2016. ISSN 0952-1976. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0952197616000324>>. Citado na página 6.
- HORVÁTH, T.; MANTOVANI, R. G.; CARVALHO, A. C. P. L. F. de. Effects of random sampling on svm hyper-parameter tuning. In: MADUREIRA, A. M. et al. (Ed.). *Intelligent Systems Design and Applications*. Cham: Springer International Publishing, 2017. p. 268–278. ISBN 978-3-319-53480-0. Citado na página 25.
- KINGMA, D. P.; BA, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. Citado 5 vezes nas páginas 19, 20, 31, 33 e 34.
- MAATEN, L. v. d.; HINTON, G. Visualizing data using t-sne. *Journal of machine learning research*, v. 9, n. Nov, p. 2579–2605, 2008. Citado 2 vezes nas páginas 46 e 48.
- MICHAEL, N. *Artificial intelligence a guide to intelligent systems*. [S.l.]: Addison Wesley, 2005. Citado 2 vezes nas páginas 9 e 16.
- MINAYO, M. C. d. S.; ASSIS, S. G. d.; OLIVEIRA, R. V. C. d. Impacto das atividades profissionais na saúde física e mental dos policiais civis e militares do rio de janeiro (rj, brasil). *Ciência & Saúde Coletiva*, SciELO Public Health, v. 16, p. 2199–2209, 2011. Citado na página 4.
- NICOLAS, C. et al. To better understand the causes of presenteeism: a systematic analysis based on the requirements of the resource requirements model. *CANADIAN PSYCHOLOGY-PSYCHOLOGIE CANADIENNE*, CANADIAN PSYCHOLOGICAL ASSOC 141 LAURIER AVE WEST, STE 702, OTTAWA, ONTARIO K1P 5J3, CANADA, v. 59, n. 3, p. 272–282, 2018. Citado 2 vezes nas páginas 1 e 2.
- OLAH, C. *Understanding LSTM Networks*. 2015. Disponível em: <<https://colah.github.io/posts/2015-08-Understanding-LSTMs/>>. Citado 2 vezes nas páginas 23 e 24.
- PANG-NING, T.; STEINBACH, M.; KUMAR, V. *Introdução ao “Data Mining”-Mineração de Dados*. [S.l.: s.n.], 2009. Citado 4 vezes nas páginas 21, 25, 32 e 53.
- PEDREGOSA, F. et al. Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011. Citado na página 26.
- PHUONG, T. M.; LIN, Z.; ALTMAN, R. B. Choosing snps using feature selection. In: IEEE. *2005 IEEE Computational Systems Bioinformatics Conference (CSB’05)*. [S.l.], 2005. p. 301–309. Citado 2 vezes nas páginas 48 e 49.

SILVA, L. N. de C. Análise e síntese de estratégias de aprendizado para redes neurais artificiais. Universidade Estadual de Campinas, 1998. Disponível em: <[ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/lnunes\\_mest/cap2.pdf](ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/theses/lnunes_mest/cap2.pdf)>. Acesso em: 19 abr. 2019. Citado 2 vezes nas páginas 13 e 21.

SILVA, R. *Vozes e sentidos do trabalho do/as operadores/as de segurança pública do Estado de Alagoas*. Maceió: Universidade Federal de Alagoas, 2015. Citado 2 vezes nas páginas 2 e 3.

# Apêndices

# APÊNDICE A – Repositório Github

Repositório do trabalho no GitHub:

<https://github.com/junioiredival/absenteismo>

Conteúdo:

- Arquivos CSV;
- Backup do banco de dados Postgres;
- Backup do esquema Pesquisa;
- Arquivos para Jupyter Notebook;
- Scripts dos experimentos;
- Imagens dos gráficos, etc.

Repositório privado, favor solicitar acesso ao autor pelo e-mail: [edivallima@ic.ufal.br](mailto:edivallima@ic.ufal.br)