

**JORGE LUIS ZANGUETTIN**

**UTILIZAÇÃO DE ALGORITMOS DE MACHINE LEARNING  
PARA CLASSIFICAÇÃO DE LICITAÇÕES DO AGRONEGÓCIO**

**TÍTULO:** Utilização de algoritmos de Machine Learning para classificação de licitações do agronegócio.

**ORIENTADOR:** João Ricardo Favan, professor das disciplinas de Internet das coisas, Data Mining e Machine Learning na Fatec Pompeia. Possui doutorado em ciências florestais, atuando principalmente na área de inteligências artificial aplicados em problemas na área de fitopatologia florestal.

**AUTOR:** Jorge Luis Zanguettin.

**TITULAÇÃO:** Tecnólogo em Big Data no Agronegócio pela Faculdade de Tecnologia Shunji Nishimura de Pompéia/SP – FATEC.

**CARGO:** Desenvolvedor Python voltado para a Ciência de Dados.

**INSTITUIÇÃO DE ENSINO SUPERIOR A QUE O AUTOR SEJA VINCULADO:** Faculdade de Tecnologia Shunji Nishimura de Pompéia/SP – FATEC.

**DEPARTAMENTO:** Tecnológico.

**ÁREAS DE INTERESSE PARA PUBLICAÇÃO:** Ciências da computação, Tecnologia da Informação, Big Data.

**ENDEREÇO PARA CORRESPONDÊNCIA:** Rua Alexandre Deboleta, 138, Flandria, Pompeia – São Paulo.

**EMAIL:** [jorgelzbr@gmail.com](mailto:jorgelzbr@gmail.com)

**TELEFONE:** (14) 99744-1842 / (14)3452-2143

**FINANCIAMENTO:** Não houve qualquer financiamento externo, todos os recursos foram próprios.

**FORMATO DE ARTIGO:** (X) Artigo Original; ( ) Artigo de Revisão; ( ) Ensaio Teórico; ( ) Revisão Integrativa; ( ) Estado da Arte; ( ) Revisão Bibliométrica; ( ) Resenha, ( ) Resumos; ( ) Entrevista; ( ) Comunicação; ( ) Dissertação; ( ) Tese.

**ANO DE DEFESA:** 2020

**PÁGINAS:** 13

**CONFLITO DE INTERESSES:** Não há conflito de interesses na produção deste estudo.

## **AGRADECIMENTO**

Agradeço todo o corpo docente da instituição que representamos, em especial ao meu orientador, por fornecer conhecimento suficiente, tanto acadêmico quanto pessoal, para o desenvolvimento do presente trabalho.

## RESUMO

O processo licitatório no Brasil tem diversas modalidades, dentre elas, o pregão eletrônico, onde existe uma grande variedade de plataformas cada uma com sua singularidade. Neste trabalho foi apresentado a utilização dos algoritmos de Regressão Logística (RL) e Rede Neural Artificial (RN) para a classificação das licitações presentes na plataforma BBMNET entre pertencentes ao agronegócio e não pertencentes ao agronegócio. Dando suporte aos algoritmos classificatórios, foi criado um dicionário de palavras e sua existência foi verificada nas licitações coletados por meio de um WEB CRAWLER. Os resultados mostram que os algoritmos de RL e RN obtiveram uma acurácia de 84% e 72% respectivamente. O algoritmo de RN obteve maior dificuldade em classificar as licitações corretamente. O modelo escolhido foi capaz de classificar as licitações entre pertencentes ou não ao agronegócio.

**Palavras-chave:** Regressão Logística; Processo Licitatório; Agricultura.

### 1. INTRODUÇÃO

O processo licitatório foi introduzido no Brasil pelo Decreto Nº 2.926, de 14 de maio de 1862 e consiste em um procedimento administrativo obrigatório para toda e qualquer aquisição de produtos ou contratação de serviços, antecedendo a assinatura de contratos com a Administração Pública (BRASIL, 1862). Previstos no artigo 45 da Lei Nº8.666/93, os tipos de licitação são três, “técnica e preço”, “melhor técnica” e “menor preço” (BRASIL, 1993).

No processo licitatório, as modalidades de licitação existentes são Concorrência, Tomada de preços, Convite ou Carta-Convite, Leilão, Concurso e Pregão presencial ou eletrônico (BRASIL, 1993). O Pregão eletrônico previsto na Lei Federal Nº10.520/2002, sendo a modalidade de licitações mais recente, é própria para a contratação de bens e serviços não determinando limite para valores, com sua disputa ocorrendo durante uma sessão pública (BRASIL, 2002).

A BBMNET é a plataforma oficial da Bolsa Brasileira de Mercadorias, desenvolvida, dentre outras finalidades, para atender a aquisição de bens e serviços de órgãos públicos através da modalidade de pregão eletrônico, concentrando as atividades para esse fim, desde a publicação dos editais de licitação até a negociação das mesmas, sendo reconhecida como uma das plataformas oficiais do Brasil através do Art. 2º da Lei Nº10.520/2002 (BRASIL, 2002).

Há uma grande quantidade de plataformas com diferentes níveis de complexidade para encontrar licitações de produtos específicos, podendo cada uma possuir sua individualidade na forma de apresentar suas licitações, itens e lotes.

WEB CRAWLER é um programa computacional que tem como objetivo analisar o código de páginas da internet para a busca de informações e usá-las para gerar listas, monitorar informações ou classificar os dados encontrados (MACHADO *et al.*, 2016).

*Machine Learning* (ML) ou aprendizado de máquina é uma área da inteligência artificial que tem como objetivo de construir algoritmos que consigam aprender de maneira autônoma, sendo capaz de reconhecer e extrair padrões de grandes volumes de dados (FOOTE, 2019).

Os tipos de aprendizado de máquina são quatro: aprendizado supervisionado, onde são fornecidos dados de treinamento que contenham as soluções desejadas (rótulos) ao algoritmo; aprendizado não supervisionado, onde os dados de treinamento entregues ao algoritmo não são rotulados; aprendizado semi-supervisionado, onde alguns algoritmos lidam com dados de tratamento parcialmente rotulados, sendo geralmente a grande parte dos dados não rotulados e uma pequena parte dos dados rotulados; aprendizado por reforço, onde o algoritmo (ou agente, neste contexto) pode observar o ambiente e realizar ações obtendo recompensas ou penalidades em troca, o objetivo deste tipo de aprendizado é o agente aprender por si só qual a melhor estratégia (GERON, 2019).

O algoritmo Regressão Logística (RL) (COX, 1958) trabalha com problemas e questões de classificação, sendo um dos mais conhecidos e utilizados no mundo em diferentes áreas, ele utiliza probabilidade para ajustar o modelo, calculando em cada amostra a probabilidade de pertencer à uma determinada classe (JAMES *et al.*, 2013).

As Redes Neurais Artificiais (RN) são técnicas computacionais que compõem um modelo matemático tendo como inspiração a estrutura neural de organismos inteligentes, caracterizando-se com a obtenção de conhecimento através de experiência (HAYKIN, 2007).

Os algoritmos utilizaram como característica para classificar licitações pertencentes ao agronegócio aquelas licitações que possuem itens e lotes voltados

para a área de alimentação ou agricultura. Carnes, frutas, verduras, hortifruti, máquinas agrícolas e grãos são exemplos de itens considerados pertencentes ao agronegócio.

O presente trabalho teve por objetivo a construção de um WEB CRAWLER para a coleta de licitações presentes na plataforma BBMNET e a utilização de algoritmos de ML para classificação de licitações entre pertencentes ao agronegócio e não pertencentes ao agronegócio.

## **2. MATERIAIS E MÉTODOS**

Foi desenvolvido um WEB CRAWLER com a linguagem PYTHON (PSF, 2020) para buscar as licitações na plataforma BBMNET com as bibliotecas SELENIUM versão 3.141.0 (MUTHUKADAN, 2018) e BEAUTIFULSOUP4 versão 4.5.3 (RICHARDSON, 2020). Este WEB CRAWLER foi executado no período de 2 dias, de 11 a 12 de maio de 2020 e coletou o total de 1.384 licitações.

As licitações coletadas foram organizadas e armazenadas no banco de dados MONGODB versão 4.2 (MONGODB, 2020), que pode ser utilizado como banco de dados não relacional, por meio de documentos do tipo JSON (HARBAND, 2020), hospedado no servidor de computação em nuvem da MICROSOFT AZURE (AZURE, 2020). Três bases de dados foram criadas, sendo elas “contas”, “licitações” e “estatísticas”, respectivamente utilizados para armazenar os usuários cadastrados na plataforma, dados sobre as licitações e as estatísticas quantitativas da base de dados.

A plataforma MOBILE foi desenvolvida utilizando o FRAMEWORK OPEN SOURCE para desenvolvimento de aplicações híbridas IONIC versão 5.4.16 (IONIC, 2020), onde foi desenvolvida a interface gráfica com o propósito de exibir as licitações classificadas entre pertencentes ou não ao agronegócio. Sendo desenvolvidas as telas *Login*, Registrar-se, Página Inicial, Licitações Agronegócio, Todas as Licitações e Configurações. A API para a conexão da plataforma MOBILE com o banco de dados foi desenvolvida utilizando o FRAMEWORK NODEJS versão 10.15.3 (NODE.JS, 2020) com as bibliotecas NODE EXPRESS versão 4.16.0 (STRONGLOOP, 2017) e MONGOOSE versão 5.9.19 (LEARNBOOST, 2011).

Com as licitações armazenadas no banco de dados, foram classificadas manualmente 100 licitações pertencentes ao agronegócio. Essas foram submetidas a uma contagem de palavras, onde foi obtido uma lista com 2.000 palavras excluindo-se termos referentes à artigos, preposições, pronomes e advérbios. Dessas palavras, foram manualmente removidos termos que não faziam referência ao agronegócio segundo o entendimento dos presentes autores. Desta forma, obteve-se um dicionário de palavras com 50 verbetes.

O dicionário de palavras utilizado no presente estudo contempla os seguintes verbetes: roçada, capinagem, roçagem, terreno, ecológica, poda, arbusto, ração, cães, gato, equino, alimentação, veterinário, café, agrícola, glifosato, herbicida, feijão, fígado, filé, muda, planta, leite, insumo, piscicultura, polpa, folhas, fruto, bife, carne, bovina, semente, vegetais, ervilha, alimento, trator, roçadeira, galinha, fruta, árvore, biológico, agrícola, pecuária, agropecuária, grão, aves, agricultura, inseticida, suíno, alimentício.

Um conjunto de 200 licitações, sendo 100 pertencentes ao agronegócio e 100 não pertencentes, foram submetidas a verificação da presença dos termos contidos no dicionário de palavras, sendo atribuído 0 a não existência de cada termo e 1 a existência do mesmo, formando assim, o conjunto de amostras que representa as licitações coletadas. Finalmente, foi adicionado a cada amostra sua categoria, sendo atribuído o valor 1 para licitações pertencentes ao agronegócio e o valor 0 para as demais licitações. Este conjunto de dados foi utilizado para treinamento e teste dos algoritmos de ML, sendo o conjunto de treinamento composto por 150 amostras e o conjunto de teste composto por 50 amostras com seleção aleatória das amostras.

Os algoritmos RL e RN foram implementados com a linguagem de programação PYTHON através da biblioteca SCIKIT-LEARN (PEDREGOSA *et al.*, 2011) que possui diversos algoritmos de classificação, regressão, pré-processamento entre outros para a utilização em trabalhos de aprendizado de máquina e mineração de dados (BROWNLEE, 2014). A implementação do algoritmo de RN foi feita com 50 neurônios na camada de entrada, 14 neurônios na camada intermediária, definido pelo método Fletcher-Goss (FLETCHER; GOSS, 1993), e 1 neurônio na camada de saída, configurados com função de ativação linear e

permitindo 100 iterações no processo de aprendizagem, os demais parâmetros foram mantidos com o valor padrão da biblioteca.

Os modelos gerados pelos algoritmos foram qualificados pela sua acurácia, sendo que o modelo com maior acurácia teve seu ajuste avaliado pelo teste de Qui-Quadrado a 95% de probabilidade (HAUCK, 1977). Este teste tem por objetivo aferir o grau de importância de cada coeficiente pertencente a um termo do dicionário de palavras, podendo verificar se cada um dos coeficientes estimados pelo modelo de classificação é significativamente diferente de zero, ou seja, se o mesmo tem significância para a classe “agronegócio”. Também foi realizada a análise da área abaixo da curva ROC (HOSMER; LEMESHOW, 2004), que tem como objetivo avaliar o desempenho de um modelo, podendo ser implementado através de um gráfico simples e robusto, apresentando a variação da sensibilidade e especificidade para os diferentes pontos de quebra (CABRAL, 2013).

### **3. RESULTADOS E DISCUSSÃO**

Os modelos de RL e RN foram testados e seus resultados foram apresentados com as métricas de acurácia, precisão, Revocação, Medida-F, assim como sua análise pelo Qui-Quadrado e a curva ROC para o modelo com maior acurácia, apresentando seu desempenho na tarefa de classificação das licitações.

A matriz de confusão tem como objetivo apresentar uma medida efetiva do modelo de classificação, mostrando o número de classificações corretas *versus* classificações preditas pelo modelo para cada classe (JOEW, 2018). A quantidade de classificações preditas corretamente pelo modelo de classificação são apresentadas nos quadrantes superior esquerdo (verdadeiros positivos) e inferior direito (verdadeiros negativos) e as classificações preditas incorretamente nos quadrantes inferior esquerdo (falsos positivos) e superior direito (falsos negativos) da matriz de confusão. Os valores falsos positivos e falsos negativos representam para o presente trabalho, além do número total de classificações incorretas do modelo, a classe na qual o modelo tende a ter maior dificuldade para realizar a classificação correta. Na Figura 1 são apresentadas as matrizes de confusão do modelo de RL (A) e modelo de RN (B).

**Figura 1 - Matrizes de confusão dos modelos de classificação RL (A) e RN (B), contendo a quantidade de classificações preditas correta e incorretamente por cada um dos modelos de classificação.**

		Classe Predita	
		Agronegócio	Não-Agronegócio
Classe Atual	Agronegócio	29	1
	Não-Agronegócio	7	13

**A**

		Classe Predita	
		Agronegócio	Não-Agronegócio
Classe Atual	Agronegócio	21	4
	Não-Agronegócio	10	15

**B**

Fonte: Elaborado pelo Autor

Foi observado que ambos os modelos de classificação tendem a ser menos precisos ao realizar classificações de licitações que não pertencem ao agronegócio, ou seja, os modelos apresentam uma maior quantidade de falsos positivos, que são licitações que não pertencem ao agronegócio classificadas como pertencentes, em vez de falsos negativos, que são licitações que pertencem ao agronegócio classificadas como não pertencentes. No contexto deste trabalho, é um cenário considerado ideal, onde o modelo classifica as licitações voltadas para o agronegócio, dentro da sua acurácia, corretamente.

Na Tabela 1, são apresentadas métricas de precisão de cada um dos modelos neste trabalho implementados. A Acurácia representa a porcentagem de acertos dos modelos com a amostra reservada para teste. A Precisão representa a capacidade dos modelos em classificar corretamente penalizado pelos falsos positivos obtidos. A Revocação representa a capacidade dos modelos em classificar corretamente penalizado pelos falsos negativos obtidos. A Medida F mostra a média harmônica entre a Precisão e o Revocação dos modelos de classificação.

**Tabela 1 - Métricas de Acurácia, Precisão, Revocação e Medida F para cada modelo de classificação.**

Modelo	Acurácia	Precisão	Revocação	Medida F
RL	0.84	0.92	0.65	0.76
RN	0.72	0.78	0.60	0.68

Fonte: Elaborado pelo Autor

Após os testes apresentados, constatou-se que o modelo de classificação RL obteve uma melhor acurácia em relação ao modelo de classificação RN para o presente estudo, sendo respectivamente 84% e 72%, consecutivamente as medidas Precisão, Revocação e Medida F também foram superiores no modelo de classificação RL, sendo respectivamente 0.92, 0.65 e 0.76.

Em seu trabalho que teve como tema a implementação de algoritmos de ML na análise de risco de crédito, Gonçalves (2005) realizou a comparação dentre outros algoritmos classificatórios, a RL e RN. O autor concluiu em seus testes que o algoritmo de RL obteve a acurácia que varia de 56,7% a 89.3%, sendo o resultado em sua amostra um pouco abaixo da RN, que foi o melhor modelo em seus testes, com uma acurácia que varia de 56,2% a 93,2% dependendo do cálculo utilizado.

Em outro trabalho, Santos *et al.* (2005) comparou dois algoritmos de ML, sendo eles RL e RN, ambos para a predição de Hepatite A. Em seus testes são apresentados os resultados e o autor conclui que a RN tem uma melhor acurácia em relação a RL, sendo 88% e 83% de acurácia para os respectivos modelos.

Diferentemente, o presente trabalho apresentou melhores resultados no algoritmo RL, onde o modelo de classificação obteve uma acurácia de 84%, sendo um resultado próximo aos demais trabalhos apresentados, porém, foi superior ao algoritmo RN, que obteve uma acurácia de 72%. Através dos resultados e da metodologia de qualificação dos algoritmos apresentados no presente trabalho, o algoritmo RL foi designado para a classificação das demais licitações coletadas e armazenadas no banco de dados.

Os estudos feitos por Gonçalves (2005) e Santos *et al.* (2005) apresentam o direcionamento para a escolha e utilização de um modelo de ML para a resolução de

um determinado problema. Da mesma forma, o presente trabalho indica o algoritmo de RL para resolução do problema de classificação estudado.

Na Tabela 2 são apresentados os atributos (verbetes contidos no dicionário de palavras) com maior significância para a classificação de licitações pertencentes ao agronegócio, contendo o seu coeficiente na equação logística, sua medida de Qui-Quadrado e sua estatística do Valor-P para o modelo de classificação RL, observada a uma significância de 95% de probabilidade.

**Tabela 2 - Coeficiente, Qui-Quadrado e Valor-P dos cinco atributos mais significantes para a classificação de licitações pertencentes ao agronegócio com o modelo de classificação RL.**

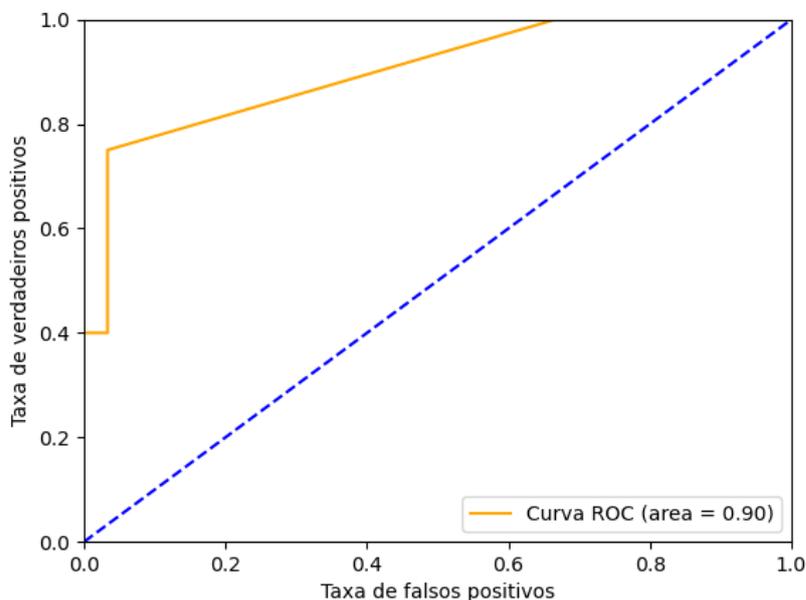
Atributo	Coeficiente	Qui-Quadrado	Valor-P
alimentício	2.47	22.75	0.000001
alimentação	1.15	9.64	0.001900
café	1.42	7.93	0.004851
feijão	0.52	7.87	0.005012
alimento	1.28	7.00	0.008150

Fonte: Elaborado pelo Autor

Na Tabela 2, foram apresentados os cinco atributos com os menores Valor-P contidos no dicionário de palavras, indicando que “alimentício”, “café”, “feijão”, “alimento” e “alimentação”, são os verbetes mais significativos para atribuir o rótulo de “pertencentes ao agronegócio” à uma licitação, quando submetida ao modelo descrito no presente trabalho.

Como outra forma de avaliação do ajuste do modelo de RL, foi utilizado a análise da área abaixo da curva ROC (*Receiver Operating Characteristic*) apresentado na Figura 2. Segundo Cabral (2013), esta curva fornece uma métrica que concede a avaliação do modelo de classificação, calculando área abaixo da curva ROC.

**Figura 2 - Curva ROC e área abaixo da curva ROC do modelo de classificação RL.**



Fonte: Elaborado pelo Autor

A curva ROC assim como a área abaixo da curva para o modelo de classificação RL apresentou um bom ajuste mostrando pouca influência dos falsos positivos e falsos negativos na classificação das licitações. A área abaixo da curva ROC resultou em um valor de 0.90, o que segundo Cabral (2013) é considerado um modelo excelente de discriminação entre licitações pertencentes e não pertencentes ao agronegócio. Após observado o ajuste do modelo de RL para classificação, este foi implementado em ambiente de produção, onde foram classificados 1.184 licitações, as quais não pertenciam nem ao conjunto de treinamento ou ao conjunto de teste, sendo elas classificadas entre pertencentes ao agronegócio (84 licitações) e não pertencentes ao agronegócio (1.100). Todas as 1.384 licitações tiveram sua classificação armazenada no banco de dados e todas foram exibidas na plataforma MOBILE, como demonstrado na Figura 3.

**Figura 3 – Plataforma MOBILE.**



Fonte: Elaborado pelo Autor

O modelo de classificação RL atingiu o propósito de classificar licitações entre pertencentes ao agronegócio e não pertencentes ao agronegócio, atingindo métricas superiores ao modelo de RN testado, demonstrando sua eficácia para finalidade do presente trabalho. Todo o código fonte produzido no presente trabalho foi disponibilizado na plataforma GitHub sob o endereço <https://github.com/JorgeZanguettin/machine-learning-licitacoes-agronegocio>.

#### **4. CONCLUSÃO**

Conclui-se que foi possível realizar a coleta de licitações provenientes da plataforma BBMNET com a utilização de um WEB CRAWLER e este se mostrou eficiente para tal finalidade. Os algoritmos de *Machine Learning* RL e RN se mostraram aptos para realizar a classificação das licitações avaliadas no presente trabalho, sendo que o algoritmo de RL demonstrou uma eficácia superior ao

algoritmo RN na classificação das licitações entre pertencentes ao agronegócio e não pertencentes ao agronegócio.

Para trabalhos futuros, sugere-se o aprimoramento da escolha das palavras que compõem o dicionário de palavras apresentado no atual trabalho, levando em consideração o estudo prévio da relevância das palavras e a sua significância para o modelo de classificação.

## 5. REFERÊNCIAS

AZURE. **Serviços de Computação em Nuvem | Microsoft Azure**, 2020. Disponível em: <https://azure.microsoft.com/pt-br/>. Acesso em: 20 maio 2020.

BRASIL. DECRETO Nº 2.926, de 14 de maio de 1862. **Aprova o Regulamento para as arrematações dos serviços a cargo do Ministério da Agricultura, Commercio e Obras Públicas**. Brasília, DF. Disponível em: <https://www2.camara.leg.br/legin/fed/decret/1824-1899/decreto-2926-14-maio-1862-555553-publicacaooriginal-74857-pe.html>. Acesso em: 17 jun. 2020.

BRASIL. Lei Nº 10.520, de 17 de junho de 2002. **Aprova o Regulamento para as arrematações dos serviços a cargo do Ministério da Agricultura, Commercio e Obras Públicas**. Brasília, DF. Disponível em: <https://www2.camara.leg.br/legin/fed/decret/1824-1899/decreto-2926-14-maio-1862-555553-publicacaooriginal-74857-pe.html>. Acesso em: 17 jun. 2020.

BRASIL. Lei nº 8.666, de 21 de junho de 1993. **Institui normas para licitações e contratos da Administração Pública**. Brasília, DF. Disponível em: [http://www.planalto.gov.br/ccivil\\_03/leis/l8666cons.htm](http://www.planalto.gov.br/ccivil_03/leis/l8666cons.htm). Acesso em: 17 jun. 2020.

BROWNLEE, Jason. **A Gentle Introduction to Scikit-Learn Machine Learning Mastery**, 15 abr. 2014. Disponível em: <https://machinelearningmastery.com/a-gentle-introduction-to-scikit-learn-a-python-machine-learning-library/>. Acesso em: 17 jun. 2020

CABRAL, Cleidy Isolete Silva. **Aplicação do modelo de regressão logística num estudo de mercado**. p. 22 e 31, 2013.

COX, David Roxbee. The Regression Analysis of Binary Sequences. **Journal of the Royal Statistical Society. Series B (Methodological)**, v. 20, n. 2, p. 215–242, 1958.

FLETCHER, Desmond; GOSS, Ernie. **Forecasting with neural networks: An application using bankruptcy data**. Information & Management, v. 24, n. 3, p. 159–167, 1 mar. 1993.

FOOTE, Keith D. **A Brief History of Machine Learning DATAVERSITY**, 26 mar. 2019. Disponível em: <https://www.dataversity.net/a-brief-history-of-machine-learning/>. Acesso em: 17 jun. 2020

GERON, Aurelien. **Mãos à Obra: Aprendizado de Máquina com Scikit-Learn & TensorFlow**. Rio de Janeiro: Alta Books, 2019.

GONÇALVES, Eric Bacconi. **Análise de risco de crédito com o uso de modelos de regressão logística, redes neurais e algoritmos genéticos**. Mestrado em Administração - São Paulo: Universidade de São Paulo, 29 jul. 2005.

HAUCK, Walter W.; DONNER, Allan. **Wald's Test as Applied to Hypotheses in Logit Analysis**. *Journal of the American Statistical Association*, v. 72, n. 360a, p. 851–853, 1 dez. 1977.

HARBAND, Jordan. **ECMAScript® 2020 Language Specification**. p. 860, 2020. Disponível em: <http://www.ecma-international.org/publications/files/ECMA-ST/ECMA-262.pdf>. Acesso em: 8 jul. 2020.

HAYKIN, Simon. **Redes Neurais: Princípios e Prática**. São Paulo: Bookman Editora, 2007.

IONIC. **Ionic Framework - Ionic Documentation**. Disponível em: <https://ionicframework.com/docs/>. Acesso em: 16 jun. 2020.

JAMES, Gareth *et al.* **An Introduction to Statistical Learning**. New York, NY: Springer New York, 2013. v. 103

JOEW. **A brief history of the confusion matrix – Joe's Data Analytics Blog**, 2018. Disponível em: <http://joew.dbsdatapoints.com/2018/06/01/a-brief-history-of-the-confusion-matrix/>. Acesso em: 17 jun. 2020

HOSMER, David W.; LEMESHOW, Stanley. **Applied Logistic Regression**. John Wiley & Sons, 2004.

LEARNBOOST. **Mongoose ODM v5.9.19**, 2011. Documentation. Disponível em: <https://mongoosejs.com/>. Acesso em: 20 maio 2020.

MACHADO, Cristian Cleder *et al.* Um Web Crawler para Projeções e Análise de Vulnerabilidades de Segurança e Consistência Estrutural de Páginas Web. **Revista de Empreendedorismo, Inovação e Tecnologia**, v. 2, n. 2, p. 3–12, 14 mar. 2016.

MONGODB. **The most popular database for modern apps**, 2020. Disponível em: <https://www.mongodb.com>. Acesso em: 20 maio 2020.

MUTHUKADAN, Baiju. **Selenium with Python**, 2018. Documentation. Disponível em: <https://selenium-python.readthedocs.io>. Acesso em: 20 maio 2020.

NODE.JS. **Node.js**. Disponível em: <https://nodejs.org/en/>. Acesso em: 16 jun. 2020.

PEDREGOSA, Fabian *et al.* **Scikit-learn**: Machine Learning in Python. *Journal of Machine Learning Research*, v. 12, p. 2825–2830, 2011.

PSF. **Python Software Foundation**. Homepage. 2020. Disponível em: <https://www.python.org/psf/>. Acesso em: 8 jul. 2020.

RICHARDSON. **Beautiful Soup Documentation**, 2020. Disponível em: <https://www.crummy.com/software/BeautifulSoup/bs4/doc/>. Acesso em: 20 maio 2020.

SANTOS, Alcione Miranda dos *et al.* Usando redes neurais artificiais e regressão logística na predição da Hepatite A. **Revista Brasileira de Epidemiologia**, v. 8, n. 2, p. 117–126, jun. 2005.

STRONGLOOP. **Express - framework de aplicativo da web Node.js**, 2020. Disponível em: <https://expressjs.com/pt-br/>. Acesso em: 20 maio 2020.