

FICHA TÉCNICA

Revista Ubiquidade, ISSN 2236-9031, v. 2, n. 2, jul./dez. 2019

Capa: Larissa Conelheiro Kovelis

Editoração e Diagramação: Prof. Dr. Juliano Schimiguel

Editora: UNIANCHIETA

Prof. Me. João Antonio de Vasconcellos

Diretor Acadêmico

Prof. Me. Anselmo de Araújo Couto

**Coordenador dos Cursos de Análise e Desenvolvimento de Sistemas,
Ciência da Computação e Sistemas de Informação**

Prof. Dr. Juliano Schimiguel

Coordenador/Editor da Revista Ubiquidade

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. É permitida a reprodução e distribuição desta obra, desde que para fins educacionais e integralmente mantidas as informações autorais. É vedado seu uso comercial, sem prévia autorização, por escrito, dos autores e da Editora.

REVISTA UBIQUIDADE

data de publicação Dezembro/2019

Copyright © 2019 UniAnchieta

EXPEDIENTE

A revista Ubiquidade é uma publicação semestral vinculada ao Curso de Bacharelado em Ciência da Computação do UniAnchieta, exclusivamente eletrônica, que pretende divulgar contribuições originais, teóricas ou empíricas, relacionadas às áreas de Tecnologia de Informação e Comunicação (TIC) e está aberta para trabalhos científicos de pesquisadores nacionais ou internacionais.

O envio de trabalhos para apreciação, assim como o pedido de informações, pode ser feito por meio do endereço: ubiquidade@anchieta.br

EDITOR

Prof. Dr. Juliano Schimiguel (UniAnchieta, Unicsul)

CONSELHO EDITORIAL

Prof.a Dra. Aline Brum Loreto, Universidade Federal de Santa Maria-Campus Cachoeira do Sul (UFSM-CS)/RS

Prof. Dr. Carlos Adriano Martins, Unicid - Universidade Cidade de São Paulo, São Paulo/SP

Prof. Dr. Carlos Eduardo Câmara (Centro Universitário Padre Anchieta, Jundiaí/SP)

Profa Dra. Cecília Sosa Arias Peixoto, Instituto Paulista de Ensino e Pesquisa, Campinas/SP

Prof. Dr. Cristiano Monteiro da Silva, Centro Universitário Anchieta, Jundiaí/SP

Prof. Dr. Hélio Rosetti Júnior, Instituto Federal do Espírito Santo, Vitória/ES

Profa. Dra. Jane Garcia de Carvalho, Unicid - Universidade Cidade de São Paulo, São Paulo/SP

Prof. Dr. Josney Freitas Silva, UEMG - Universidade do Estado de Minas Gerais - UEMG, Frutal/MG

Prof. Dr. Juliano Schimiguel (UniAnchieta, Unicsul)

Prof. Me. Juliano Silva Marçal (Centro Universitário Padre Anchieta, Jundiaí/SP)

Prof. Dr. Luciano Soares Pedroso, Universidade Federal dos Vales do Jequitinhonha e Mucuri - UFVJM, Teófilo Otoni/MG

Profa Dra. Lucy Mirian Campos Tavares Nascimento, Instituto Federal de Educação, Ciência e Tecnologia de Goiás, Formosa/GO

Prof. Dr. Marcelo Eloy Fernandes, Universidade Nove de Julho, São Paulo/SP

Profa. Ma. Nadia Vilela Pereira, IFTO — Instituto Federal do Tocantins, Campus Palmas
Prof. Dr. Vivaldo José Breternitz, Universidade Presbiteriana Mackenzie, São Paulo/SP
Profa Dra. Viviane Sartori, Universidad Europea del Atlántico (Uneatlatico), Salamanca,
Espanha

Todos os direitos reservados e protegidos pela Lei 9.610 de 19/02/1998. É permitida a reprodução e distribuição desta obra, desde que para fins educacionais e integralmente mantidas as informações autorais. É vedado seu uso comercial, sem prévia autorização, por escrito, dos autores e da Editora.

SUMÁRIO

PREFÁCIO.....	5
UM SISTEMA DE GERENCIAMENTO DE PROCESSOS DE INFORMAÇÃO EM MÓDULO TECNOLÓGICO DE COLETA DE DADOS EM CAMPOS DE PRODUÇÃO DE SEMENTES DE MILHO (Josiane Rezende, Cecília Sosa Arias Peixoto)	6
A IMPLEMENTAÇÃO DE UMA FERRAMENTA DE <i>BUSINESS INTELLIGENCE</i> EM UMA EMPRESA DE PEQUENO PORTE (Vivaldo José Breternitz).....	19
ANÁLISE DE UM ALGORITMO DESENVOLVIDO PARA SOLUÇÃO DE DEEP LEARNING UTILIZANDO REDES NEURAIS CONVOLUCIONAIS PARA DIFERENCIAR GATOS DE CACHORROS (Jonas de Souza Tega, Carlos Eduardo Câmara)	29
APLICAÇÃO E ANÁLISE DO MÉTODO ÁGIL SCRUM EM UM PROJETO DE “M.E.S [MANUFACTURING EXECUTION SYSTEM]” (Mário Vinícius Vieira, Cecília Sosa Arias Peixoto)	64
IMPLEMENTANDO UMA PLATAFORMA BIG DATA PARA VISUALIZAÇÃO DE DADOS GERADOS POR DISPOSITIVO IOT (Jair Campos de Oliveira Júnior, Juliano Schimiguel)	85

PREFÁCIO

Este volume 3, número 2, da Revista Ubiquidade, traz publicações de importantes pesquisadores nacionais e internacionais. Podemos citar pesquisadores do Instituto Presbiteriano Mackenzie, da Instituição Educacional de Piracicaba (PECEGE), do Centro Universitário Anchieta (UNIANCHIETA), e da Universidade Cruzeiro do Sul.

Rezende e Peixoto tiveram como objetivo mostrar a aplicabilidade da implementação de um sistema tecnológico de gerenciamento de dados em campos de produção agrícola de milho. Os autores tiveram como premissa analisar quatro aspectos que uma implementação tecnológica no meio rural traz, sendo eles: gerenciamento de informações, controle e monitoramento, investimento tecnológico e praticidade em realizar atividades rotineiras.

Breternitz relata em seu trabalho o processo que levou à implementação de ferramenta de Business Intelligence em uma empresa comercial de pequeno porte que opera como uma trading company, intermediária entre fabricantes e consumidores ou distribuidoras, em operações de importação de produtos químicos e resinas plásticas.

Tega e Câmara trazem os conceitos básicos sobre Redes Neurais Convolucionais, e como a sua fusão com Aprendizado Profundo (Deep Learning) se encaixam para o reconhecimento de imagens.

Vieira e Peixoto apresentam a aplicação da metodologia de gestão de projetos, denominada Scrum, em uma empresa multinacional do setor de tecnologia da informação. Segundo os autores, essa empresa é prestadora de serviço e atua no ramo automobilístico, onde existe grande demanda de projetos de "Manufacturing Execution System" (MES), de curta duração e alto grau de criticidade. Estes projetos envolvem vários setores da fábrica e estão diretamente ligados ao ambiente de produção.

Oliveira Jr. e Schimiguel desenvolveram e apresentaram um dispositivo de IoT (Internet das Coisas) para coleta de dados gerados por sensores, e também com a implementação de uma plataforma Bigdata para extração e visualização de dados.

Prof. Dr. Juliano Schimiguel

Coordenador da Revista Ubiquidade

UM SISTEMA DE GERENCIAMENTO DE PROCESSOS DE INFORMAÇÃO EM MÓDULO TECNOLÓGICO DE COLETA DE DADOS EM CAMPOS DE PRODUÇÃO DE SEMENTES DE MILHO

AN INFORMATION PROCESS MANAGEMENT SYSTEM IN DATA COLLECTION TECHNOLOGICAL MODULE IN MAIZE SEED PRODUCTION FIELDS

Josiane REZENDE

josiane-rezende@hotmail.com

MBA em Gestão de Projetos. PECEGE.

Cecilia SOSA ARIAS PEIXOTO

cecilia.sosaarias@gmail.com

Orientadora. PECEGE.

Resumo

Este trabalho tem como objetivo mostrar a aplicabilidade da implementação de um sistema tecnológico de gerenciamento de dados em campos de produção agrícola de milho. Teve como premissa analisar quatro aspectos que uma implementação tecnológica no meio rural traz, sendo eles: gerenciamento de informações, controle e monitoramento, investimento tecnológico e praticidade em realizar atividades rotineiras. Para tanto, foi preparado um plano de avaliação de usabilidade do software implementado, baseada na escala Likert. Os resultados deste questionário mostraram que os avanços tecnológicos no setor agrícola, apesar de ainda engatinhar no Brasil, trazem excelentes benefícios às estruturas de controle da gestão agrícola

Palavras-Chave

plano de avaliação de usabilidade, avanço tecnológico, gerenciamento de informações, coleta de dados, planejamento,

Abstract

This work aims to show the applicability of implementing a technological data management system in agricultural corn production fields. The premise was to analyze four aspects that a technological implementation in the rural environment brings, namely: information management, control and monitoring, technological investment and practicality in carrying out routine activities. For that, a usability evaluation plan of the implemented software was prepared, based on the Likert scale. The results of this

questionnaire showed that technological advances in the agricultural sector, although still crawling in Brazil, bring excellent benefits to agricultural management control structures.

Keywords

usability evaluation plan, technological advances, information management, data collection, planning.

INTRODUÇÃO

Atualmente, o setor de negócios agrícolas vem se consolidando como um dos mais importantes para a economia brasileira e estima-se que represente, em média, um terço do Produto Interno Bruto [PIB]. Apesar de tão grande representatividade, observa-se que ainda existem deficiências no campo da gestão quando nos referimos aos inúmeros avanços tecnológicos de gerenciamento de sistemas de informações provenientes desses processos. Até meados dos anos 1990, os sistemas de controle de informações dos processos de trabalho, na maioria dos negócios agroindustriais, eram feitos individualmente. Cada “departamento” controlava suas informações, variáveis, resultados, etc., e as armazenava em banco de dados único ou não. Não havia, até então, padronização alguma. No entanto, quando necessários, eram compilados através de recursos de informática – planilhas em Excel e gráficos, por exemplo, e confeccionados os relatórios gerenciais que davam a possibilidade de fazer análises e enxergar as oportunidades de melhorias.

Foi assim que nasceram os historiadores de plantas industriais, que nos dias de hoje, dão inúmeras oportunidades de aperfeiçoamento dos processos industriais. Os historiadores têm como função principal integrar todas as plantas de uma indústria eliminando assim as “ilhas de informação”, permitindo uma real redução de tempo e custos na obtenção de informação do processo industrial (SEIXAS FILHO, 2004a).

Um ponto de extrema importância em todo esse processo industrial para uma análise confiável e segura é a obtenção de dados advindos da produção e os dados da gestão de negócios para então serem aplicados na ferramenta apropriada.

Hoje, no mercado, existem duas diferentes ferramentas de otimização; são elas: o PIMS (que vem do inglês “Process Information Management System”) e Manufacturing Execution System [MES] que são sistemas de gestão integrados que oferecem diferentes maneiras de disponibilizar informações sobre os processos, sejam no setor agrícola, industrial e/ou administrativo para tomada de decisão, possibilitando uma maior efetividade na gestão dos dados neles coletados e armazenados.

Process Information Management Systems [PIMS] são sistemas de coleta de informações contínuas de diferentes fontes e processos de trabalho com grande capacidade de armazenamento e, na maioria das vezes, em banco de dados único. Este Banco de dados é utilizado por diferentes ferramentas em forma de relatórios gerenciais. De acordo com Seixas Filho (2011), o PIMS nasceu na indústria de processos contínuos, mais propriamente na indústria química e petroquímica para resolver o problema da fragmentação de dados e proporcionar uma visão unificada do processo.

O PIMS passa a se constituir uma ferramenta fundamental para o engenheiro de processos. A partir de uma estação, ele pode visualizar tanto os dados de tempo real como os históricos da planta. Através de suas funcionalidades, é possível montar tabelas, gráficos de tendência e sinóticos, e eliminar as ilhas de informação, concentrando em uma única base de dados de informações todos os aspectos de uma planta. Até hoje, as indústrias química e petroquímica constituem o principal nicho para o PIMS, mas outros mercados foram rapidamente conquistados como: papel e celulose, siderurgia, cimento, mineração, enfim, todos os processos contínuos. Em um contexto mais amplo, PIMS são sistemas de aquisição de dados que, fundamentalmente, recuperam os dados do processo oriundos de diferentes fontes, os armazenam em um banco de dados único e os disponibilizam através de diversas ferramentas (CARVALHO et. al, 2005). Com o uso do PIMS, torna-se possível visualizar e organizar dados históricos em tempo real, de forma que estes dados obtenham informações úteis para a gestão do processo, seja ele industrial ou não (Figura 1) (URBANO, 2009). Em adição, um PIMS possui algumas funcionalidades como montar tabelas, gráficos e relatório dinâmicos, pois tem como viés concentrar informações, e isso possibilita uma visão ampla e consolidada de todo processo produtivo.

Na Figura 1 é possível observar que a camada inferior de um PIMS é onde acontece a coleta dos dados que, em sua maioria, são obtidos através de equipamentos ligados diretamente às fontes de informação primária. Como por exemplo, os sistemas “Supervisory Control & Data Acquisition” [SCADA] e Sistemas Digitais de Controle Distribuído [SDCD] e outros, cada um com sua características ou layout, mas sempre com a mesma funcionalidade final que é enviar os dados coletados para um único banco de armazenamento – segunda camada – na qual as informações são mantidas de forma centralizada a fim de prover alguma forma de compressão dos dados provenientes da fase anterior por cerca de até 15 anos (SEIXAS, 2014). A última camada, chamada de exibição, é responsável por reintegrar dados históricos com extrema agilidade ou em tempo real da camada de armazenamento, disponibilizando de maneira a escolher o modo visual a ser analisado, através de gráficos de tendência ou não, relatórios de diferentes características e possibilidades de correlação. Os PIMS permitem realizar comparativos da situação atual com informações históricas armazenadas e emitir uma serie de relatórios gerenciais para a tomada de decisão.

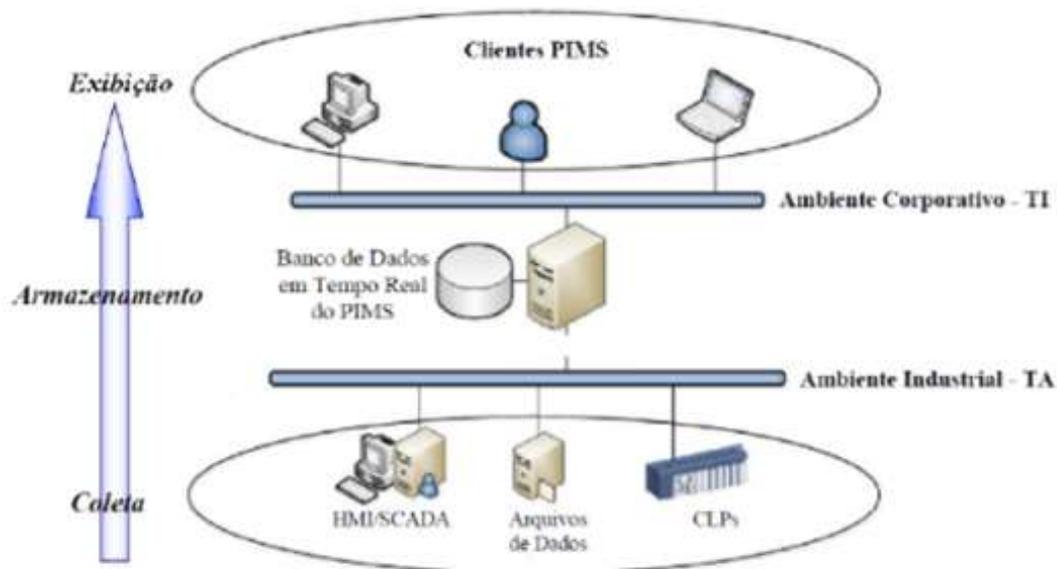


Figura 1. Arquitetura de um PIMS

Fonte: Elaborado pelo autor, com base em Urbano (2009).

A importância da coleta de dados na camada superior de um PIMS vem ganhando força inclusive devido às necessidades regulamentações junto ao governo, relacionados à produção, rastreabilidade, divulgação de indicadores de processo, análise gráficas, etc.

Tendo em vista estes quesitos, um dos principais pontos a serem observados no setor agrícola, dá-se pela necessidade em se obter um sistema de gerenciamento de informações advindos do processo de produção de campo de Sementes de Milho e junto disso vem a preocupação de como os dados devem ser coletados, armazenados e compilados. Isto representa uma preocupação gerencial. À medida que a gerência identifica as vantagens da utilização destes sistemas, os técnicos em campo resistem na utilização destas tecnologias.

Uma das características de maior impacto citadas pelos técnicos é a dificuldade de interagir com o sistema para realizar a entrada de dados em campo. Estes tipos de problemas são tratados por uma área específica da Engenharia de Software (PRESSMAN, 2006). A Engenharia de Software é a responsável pelo desenvolvimento de qualquer sistema. Segundo Pressman (2006), Engenharia de Software é um processo com métodos e ferramentas destinados à construção de software de qualidade. Dentro da Engenharia de Software, a área da Engenharia de Usabilidade (NIELSEN, 1993) é a responsável por construir um sistema fácil de ser aprendido e de usar, agradável, eficiente, fácil de lembrar sua operação e que auxilie os usuários na recuperação em caso de erro durante o trabalho com o sistema. Estes atributos citados (aprendizagem, satisfação, eficiência, memorização, e tratamento de erros) fazem parte de um conceito de qualidade definida pela primeira vez pela norma ISO 9126 e depois aperfeiçoada pela norma ISO/IEC 12119 (1994), ISO 924111 e norma de avaliação de qualidade ISO/IEC 14598, para todos os sistemas de software. Segundo a norma ISO 9126, o conceito usabilidade está formado

por um conjunto de atributos relacionados ao esforço necessário para seu uso e para o julgamento individual de tal uso por determinado conjunto de usuários (ISO/IEC 91261, 1995). A definição de usabilidade da norma ISO 924111 (1998), inclui “a capacidade de um produto poder ser usado por usuários específicos para atingir objetivos específicos com eficácia, eficiência e satisfação em um contexto específico de uso”. Considerar este conceito é fundamental para o êxito da utilização de qualquer sistema. No caso específico do PIMS, pode-se destacar que existem dois grandes grupos de usuários que têm apreciações diferentes sobre a usabilidade dos PIMS, os usuários administrativos e os de campo. Para o primeiro grupo, os PIMS apresentam inúmeras vantagens. Enquanto que aos usuários do segundo grupo, destacam desvantagens relacionadas com a usabilidade.

Este trabalho tem como objetivo apresentar como a implementação do PIMS no campo agrega valor no que tange ao cruzamento das informações para um gerenciamento mais eficiente e um maior ganho de resultados para os usuários administrativos. Este trabalho também apresenta uma pesquisa *survey* realizada para avaliar a importância da característica de qualidade da usabilidade no PIMS numa empresa de produção agrícola de milho aplicada aos usuários de campo.

MATERIAL E MÉTODOS

O estudo caracterizou-se, na primeira fase do trabalho, como uma pesquisa bibliográfica sobre as principais características do PIMS suas vantagens e desvantagens. O estudo também consta com os resultados da experiência da autora na implantação do PIMS numa empresa de produção agrícola de milho. Na segunda fase do trabalho foi feita uma avaliação de usabilidade do sistema que atua em campo.

É por meio da avaliação de usabilidade que é possível identificar se o sistema será bem aceito pelo usuário em termos de funcionalidade e de interação. Para realizar a avaliação, foi elaborado um questionário baseado nos atributos de usabilidade (NIELSEN, 1993): Habilidade de Aprendizado, Eficiência de Uso, Habilidade de Memorização, tratamento de Erros e Satisfação.

A aplicação do questionário foi feito aos técnicos agrícolas com início no segundo semestre de 2017, e término em novembro do mesmo ano.

Elaboração do questionário para Avaliação da Usabilidade

O questionário elaborado nesta pesquisa foi enviado a cada técnico agrícola (usuários) a fim de avaliar a efetividade do sistema FARMS (Figura 1).

O questionário foi elaborado em 5 seções, a saber:

- Questões relacionadas à Habilidade de Aprendizado
- Questões relacionadas à Eficiência de Uso
- Questões relacionadas à Habilidade de Memorização
- Questões relacionadas ao Tratamento de Erros
- Questões relacionadas à Satisfação do usuário em relação ao software

Instruções de Preenchimento do Questionário		
<i>Avaliar cada questão abaixo e pontuá-las de acordo com sua opinião utilizando a seguinte legenda:</i>		
(1) discordo inteiramente, (2) discordo, (3) nem concordo nem discordo (4) concordo (5) concordo inteiramente.		
1. Questões relacionadas à Habilidade de Aprendizado		Nota
Q1	Os treinamentos oferecidos previamente foram devidamente aplicados	
Q2	Suas habilidades tecnológicas facilitam o uso da ferramenta	
Q3	Você considera o FARMS um aplicativo Autoexplicativo em suas funções.	
Q4	A tradução feita no software FARMS facilitou o uso e estimulou o aprendizado ao manuseá-lo	
Q5	Na Tela Principal, a disposição das tarefas a serem completadas são sequenciais facilitando a aprendizagem do software.	
2. Questões relacionadas à Eficiência de Uso		Nota
Q6	O tamanho da tela do tablet é suficientemente adequado para o uso no campo	
Q7	A velocidade da rede 3G oferecida é suficientemente adequado para a sincronização dos dados para o SeedPro	
Q8	A não acentuação gráfica no FARMS ajuda na eficiência ao digitar dados no software	
Q9	Os dados imputados no FARMS atenderam às necessidades de monitoramento das atividades de campo da companhia	
3. Questões relacionadas à Habilidade de Memorização		Nota
Q10	Na Tela Principal, a disposição e tamanho dos botões facilitam a memorização do padrão a ser seguido ao imputar os dados	
Q11	A navegação entre uma tela e outra segue um raciocínio lógico facilitando a memorização	
4. Questões relacionadas à Tratamento de Erros		Nota
Q12	Quando inseridos dados errados a correção dos mesmos é fácil e rápido	
Q13	Os bloqueios de caracteres Alfanuméricos ajudam a simplificar a entrada dos dados	

Q14	A função Auto-correção alfa-númerica está disponível na ferramenta afim de corrigir erros de digitação	
5. Questões relacionadas à Satisfação do usuário em relação ao software		Nota
Q15	O suporte prestado em relação ao uso do software é eficiente	
Q16	A implementação do FARMS agilizou tarefas repetitivas do dia a dia que antes eram feitas manualmente em papel e caneta	
Q17	O uso de tecnologias no campo é um diferencial positivo que a cia oferece aos funcionários	
Q18	A utilização desse software pode colaborar no planejamento das atividades agrícolas.	
Q19	A operacionalidade do software no tablet não atendeu as expectativas em todas as aplicações	

Figura 1. Questionário

Fonte: Dados originais da pesquisa

A métrica utilizada no questionário é a de Likert (LIKERT, 1975). A métrica identifica a concordância às questões em uma escala do tipo: (1) discordo inteiramente, (2) discordo, (3) nem concordo nem discordo (4) concordo, (5) concordo inteiramente.

RESULTADOS

A utilização do modelo PIMS traz inúmeros benefícios e destaca-se como vantagem principal a otimização do armazenamento e a agilidade na busca de informações, evitando informações “off-line” e em ilhas.

O PIMS facilita a implantação de outros módulos de software, que integra ao chão de fábrica propiciando ainda mais recursos para as companhias, como o “Supply Chain Manager”, sistemas ERP, “Manufacturing Execution Systems” [MES], oferecendo a capacidade de gerar outros dados através de cálculos e armazená-los por longo períodos de tempo sem ter que enviá-los a um mainframe, o que constitui um grande ganho para o analista de processos, pois ele passa a gerar seus próprios relatórios sem ter que se preocupar se o dado se origina num CLP, num sistema SCADA ou SDCD e deixa de depender do departamento de informática.

Os benefícios gerados pela implementação de um sistema PIMS vão além da integração e armazenamento dos dados, e baseado em um dos artigos publicados por Seixas Filho (2014) vale a pena ressaltar alguns que são reais, mas aparecem em segundo plano:

- Padronização das informações e eliminação de interfaces manuais;
- Melhoria na consistência dos dados;
- Melhoria do fluxo da informação o que implica em agilidade na tomada de decisões;
- Democratização da informação, onde o PIMS permite que qualquer usuário tenha acessos aos dados coletados instantaneamente;

- Unificação/centralização dos dados, dessa forma qualquer computador da empresa consegue apresentar dados de diversas áreas do processo;
- Disponibilização em Tempo real de dados;
- Possibilidade de comparativos baseados em dados reais;
- Possibilidade de análises sobre o desempenho dos processos de trabalho;
- Visualização dos dados em diversas formas, como gráficos e tabelas;
- Elimina retrabalho ambiguidades, erros e redundâncias.

Dentre tantas vantagens, para área agrícola, que é área foco deste trabalho, a capacidade de realizar diversas combinações de dados geográficos, consultas por assuntos/tópicos/safras, como por exemplo: por tipo de produção, região de produção, por tipo de solo, entre outros, facilitando o entendimento do usuário; podendo inclusive chegar a um nível de detalhamento mais aguçado como filtrar por campos em que a infestação das plantas aconteceu com mais incidência ou não desde que parâmetros sejam estabelecidos para isso.

O PIMS dá possibilidades da aplicação da agricultura de precisão na lavoura, diminuindo custos, recursos, economizando equipamentos e insumos e aumentando a qualidade de toda a sua produção, seja de cana, de grãos, de cítricos, ou outras.

Para que um PIMS seja implementado em uma organização, bem como todo novo processo que tem suas vantagens, ele traz alguns pontos que podem ser considerados como desvantagens. Porém para cada uma delas vale levar em consideração o poder aquisitivo e de desenvolvimento investido.

Entre as desvantagens é possível citar:

- Alto custo de investimento tanto para recursos de TI, como para treinamento e capacitação de colaboradores;
- Custo de suporte técnico, pois requer conhecimentos para parametrização de todo o sistema, além da capacitação dos colaboradores,
- Requer interface com ERP;
- Requer um bom funcionamento de Sistema de Gestão da Qualidade estabelecido a fim de que os procedimentos dos processos sejam estabelecidos para atingir um padrão operacional de qualidade.
- Requer tempo para trabalhar a conscientização da equipe sobre a importância e a sensibilidade do sistema, e como e quanto cada papel interfere nos resultados.
- Requer pessoas qualificadas para o gerenciamento e auditoria dos dados apontados e desempenho bem como direcionamento estratégico.

Ainda que o PIMS tenha diversas vantagens, ele pode se tornar uma ferramenta inadequada se a companhia não tiver a cultura de trabalhar com fontes de dados padronizadas, como, na maioria das vezes as indústrias são. Para uma eficiente implantação de um PIMS, este fator deve ser rigorosamente levado em conta, pois esse sucesso dependerá de esforços inclusive de investimentos em uma nova infraestrutura para sua adequação.

Apesar da lista de desvantagens não se limitar em 2 ou 3 pontos, o PIMS é um sistema revolucionário para as organizações, e, se bem implementado, gerador de valores de gestão. Além disso, a coleta de dados ser realizada de forma tecnológica no campo agiliza o carregamento das informações de forma rápida e sujeitas a menos erros. Anteriormente, na coleta, os técnicos escreviam em papeis que posteriormente seriam digitados para

alimentar os diversos sistemas o que levava a uma série de erros e ambiguidades junto com uma lentidão no processo de todo o sistema.

Foi realizado um estudo de caso na Indústria de Sementes de Milho Híbrido na qual a autora de este trabalho desempenha o cargo de Analista de Manufatura e Qualidade, o sistema da coleta automatizado através de tablets, chamado FARMS que significa Field Activities Records Management Systems, foi implantado no ano de 2015. Anteriormente, no ano 2013, tinha sido implantado os módulos principais do PIMS, chamado de SeedPro, que até então era manuseado apenas por colaboradores administrativos que recebiam informações feitas pelos técnicos agrícolas em papel e transcrevia para o sistema.

Entre as funcionalidades oferecidas pelo o SeedPro é possível citar:

- Cadastrar Cooperantes;
- Cadastrar contratos;
- Cadastrar produtos químicos que usamos nos campos;
- Cadastrar Materiais a ser plantada (que é a semente de milho);
- Analisar Solo;
- Comparar e identificar regiões geográficas a fim de verificar qual a que melhor se encaixa em cada material a ser cultivado;
- Padronizar informações;
- Controlar gastos acontecendo no campo;
- Identificar áreas mais produtivas através de dados históricos;
- Acompanhar se as instruções de plantio e cultivo estão sendo seguidas;
- Armazenar dados de colheita;
- Realizar inspeções no campo, como florescimento, por exemplo;
- Acompanhar se a produtividade esperada e se está sendo alcançada ao longo da safra;
- Avaliar os melhores cooperantes;
- Avaliar os melhores híbridos;
- Extrair relatórios de todas as atividades através do Excel que permite cruzar todas as informações que foram entradas no sistema.

A informatização da coleta de dados em campo com o FARMS, através dos técnicos trouxe vários ganhos, como a diminuição de retrabalhos, e agilidade em obter as informações online. Porém, com o passar do tempo, diversas reclamações por parte dos usuários técnicos apareceram e notou-se que ajustes se faziam necessários. Sendo assim, a autora deste trabalho realizou uma avaliação de usabilidade do sistema FARMS. Além disso, a avaliação foi executada, os resultados obtidos e analisados resultando nas conclusões de este artigo.

Coleta de dados

Durante a coleta de dados, o questionário foi disponibilizado para o total de 5 técnicos para que se pudesse avaliar a usabilidade do FARMS.

Tabela 1. Número de respostas por categoria para cada variável estudada

<u>Variáveis</u>	Concordo Plenamente (5)	Concordo (4)	Indiferente (3)	Discordo (2)	Discordo Completamente (1)
FQ1	6	4	0	1	1
FQ2	4	3	0	2	3
FQ3	1	6	3	0	2
FQ4	10	0	1	1	0
FQ5	0	9	2	1	0
FQ6	6	6	0	0	0
FQ7	1	5	2	3	1
FQ8	10	0	0	2	0
FQ9	5	3	3	1	0
FQ10	3	2	4	3	0
FQ11	0	5	0	7	0
FQ12	0	4	4	2	2
FQ13	0	6	4	2	0
FQ14	0	10	2	0	0
FQ15	6	4	0	0	2
FQ16	5	5	0	0	2
FQ17	5	5	0	0	2
FQ18	8	2	1	1	0
FQ19	10	0	2	0	0

Fonte: Resultados originais da pesquisa

DISCUSSÃO

Baseado nas respostas encontradas no questionário de avaliação, foi identificado que precisa ser realizado alguns ajustes técnicos no software a fim de atingir a satisfação e expectativas dos usuários de forma mais ampla.

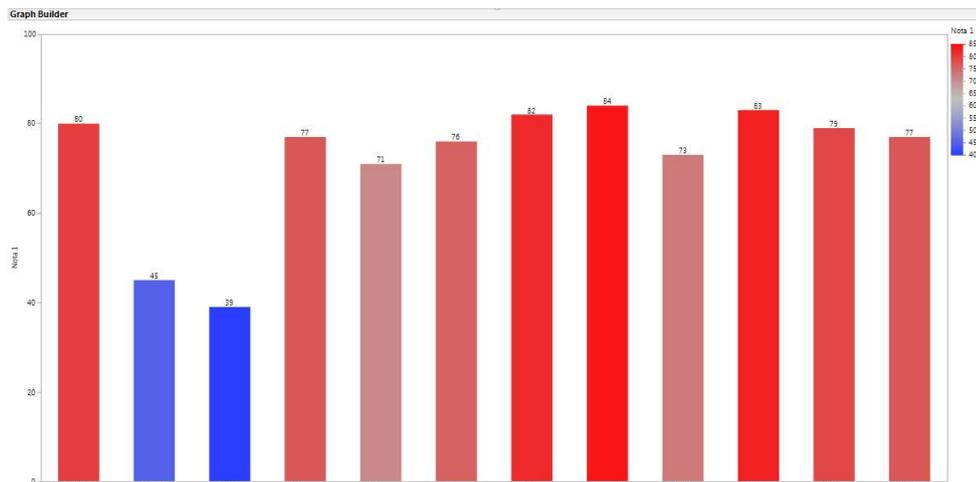


Figura 2. Respostas encontradas no questionário de avaliação

Fonte: Resultados originais da pesquisa

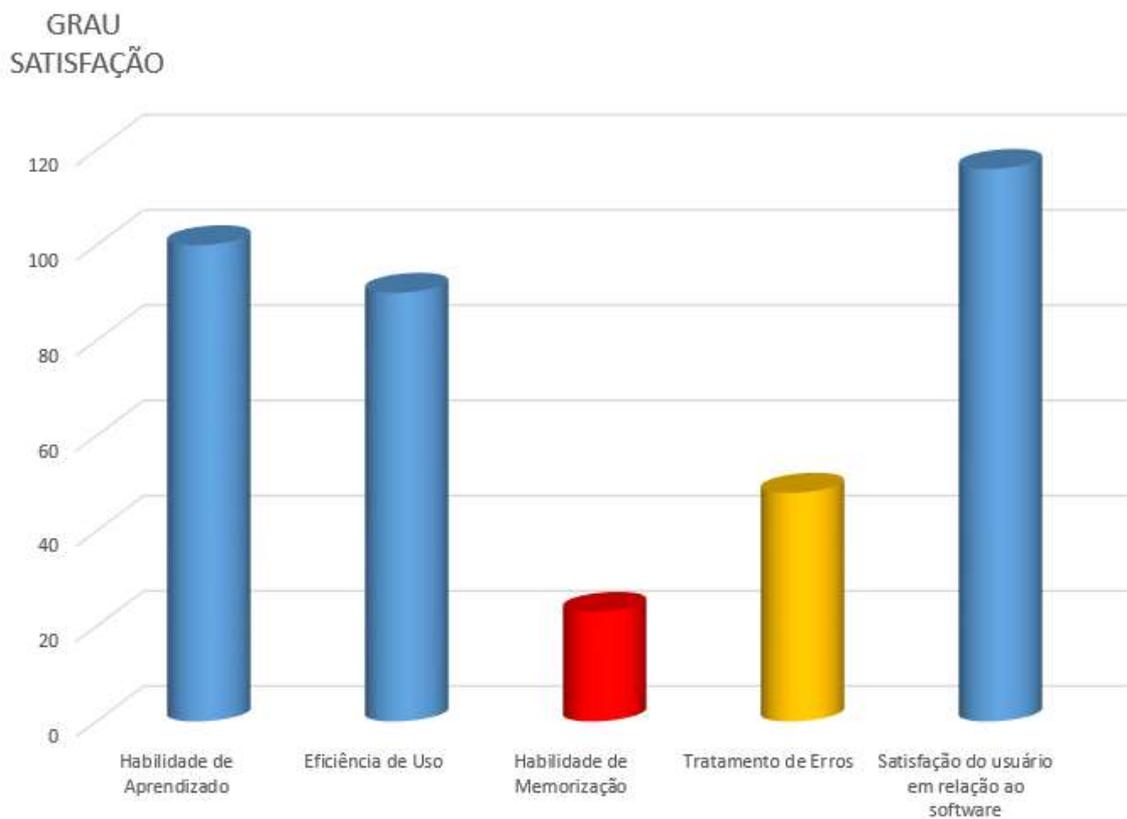


Figura 3. Respostas encontradas no questionário de avaliação

Fonte: Resultados originais da pesquisa

A Figura 2 mostra que apenas 2 funcionários tem uma visão bem diferente dos demais, dando as notas mais baixas para o processo

A Figura 3 indicou um baixo grau de satisfação nas áreas que implicam em Habilidade de Memorização e Tratativa de erros por parte dos usuários na parte de técnica da coleta dos dados. Com resultados em mãos, a autora deste trabalho realizou entrevistas pessoalmente com os usuários afim de identificar quais eram as dificuldades que eles sentiam ao responder o questionário dando nota baixa para esses quesitos.

Nesta entrevista foi possível identificar que as trocas de telas, e sequência das atividades no aplicativo poderiam facilmente ser alteradas para atender de forma mais satisfatória ao quesito 1 (Habilidade de Memorização). No quesito Tratamento de Erros faltava somente um ajuste no formato da data/calendário que também seria extremamente fácil de ser corrigida.

Apesar desses ajustes, a satisfação no uso da tecnologia ficou evidenciada desde a resposta adquirida no questionário que foi enviado anteriormente à entrevista pessoal.

CONCLUSÕES

O objetivo deste trabalho foi estudar a usabilidade do módulo tecnológico de coleta de dados em campos de produção de sementes híbridas de Milho através da implementação de um PIMS, levando-se em consideração as necessidades da companhia em estabelecer um sistema de gestão que disponibilize informações ágeis e confiáveis provenientes das atividades desempenhadas manualmente nos campos agrícolas.

Neste trabalho, evidenciou-se que a implementação de um módulo tecnológico de caráter PIMS oferece inúmeros recursos de gestão e reduz custos indiretos como retrabalho, porém deve ser avaliado o contexto organizacional a fim de garantir que o software se adequa às necessidades existentes, bem como os recursos disponíveis atendem também aos requerimentos de uso do módulo.

Concluiu-se que, apesar ajustes necessários, os usuários estão satisfeitos com o fato de ter a tecnologia no campo, e que ela facilita as atividades que eles desempenham. Mas vale a pena olhar com mais criticidade para a aplicabilidade técnica no manuseio do aplicativo.

REFERÊNCIAS BIBLIOGRÁFICAS

CARVALHO, F. B.; TORRES, B. S.; FONSECA, M. O.; SEIXAS FILHO, C.; Sistemas PIMS: conceituação, usos e benefícios. Anais do Seminário de Automação de Processos, Santos, 2003.

ISO/9241-11. Ergonomic Requirements for Office Work with Visual Display Terminals - Part 11 – Guidance on Usability. International Standards. 1998.

ISO/IEC 9126-1. International Standard. Information Technology – Software Quality Characteristics and Metrics – Part 1: Quality Characteristics and Sub Characteristics. June 1995. Geneva. Switzerland. 1995.

- ISO/IEC IS 9126-1. Software Engineering - Product Quality – Part 1: Quality Model. International Organization for Standardization, Geneva, Switzerland.
- Netto, A. 2004. IHC: Modelagem e Gerência de Interfaces com o Usuário. Visual Books, Florianópolis, SC, Brasil. 2001.
- LIKERT, R. A *Organização Humana*. São Paulo: Atlas, 1975.
- NIELSEN, J. *Usability Engineering*. Boston: Academic Press, USA. Organization, Geneva, Switzerland. 1993.
- PRESSMAN, R.. *Engenharia de Software*. 6 eds. MacGraw-Hill, São Paulo, SP, Brasil. 2006.
- SEIXAS FILHO, Constantino. (s.d.). PIMS -Process Information Management System – Uma introdução. Acesso em 27 de Outubro de 2017, Disponível em: <http://www.cpdee.ufmg.br/~seixas/PaginaII/Download/DownloadFiles/Pims.PDF>>
- SEIXAS FILHO, Constantino; SZUTER, Marcelo. Programação Concorrente em ambiente Windows – Uma visão de automação; Editora da UFMG, 1993.
- SEIXAS FILHO, Constantino. Notas de Aula – Capítulo 6 PIMS – Process Information Management System; url: <http://www.cpdee.ufmg.br/~seixas>, 2005.
- SOUZA, A. Sistema de Gerência de Informação de Processo Industriais via WEB. Tese (Mestrado em Engenharia Elétrica) – Centro de Tecnologia/UFRGN, 2005. Disponível em: <ftp://ftp.ufrn.br/pub/biblioteca/ext/btd/AlessandroJS.pdf> Acesso em: 11 set. 2017.
- URBANO, L. R. H. *Gerenciamento de ativos de automação: Uma solução baseada em PIMS*. 4o Seminário Nacional de sistemas Industriais e Automação. 2009.

A IMPLEMENTAÇÃO DE UMA FERRAMENTA DE *BUSINESS INTELLIGENCE* EM UMA EMPRESA DE PEQUENO PORTE

THE BUSINESS INTELLIGENCE TOOL IMPLEMENTATION IN A SMALL BUSINESS

Vivaldo José BRETERNITZ

vjbreternitz@mackenzie.br

Faculdade de Computação e Informática da Universidade Presbiteriana Mackenzie,
FCI/UPM, São Paulo/SP, Brasil

RESUMO

Este trabalho relata o processo que levou à implementação de ferramenta de *Business Intelligence* em uma empresa comercial de pequeno porte que opera como uma *trading company*, intermediária entre fabricantes e consumidores ou distribuidoras, em operações de importação de produtos químicos e resinas plásticas. A empresa tem a percepção de que a implementação da ferramenta incrementou seu volume de negócios. Confirmou-se a importância da utilização de *framework* para a escolha de solução adequada às necessidades da empresa e as dificuldades de organizações de pequeno porte implantarem ferramentas de *software* sem o apoio de pessoal externo. O trabalho tem como objetivo compartilhar a experiência adquirida, de forma a que esta possa ser útil a pesquisadores e profissionais envolvidos com o tema.

PALAVRAS CHAVE

Business Intelligence; empresas de pequeno porte; implantação.

ABSTRACT

This paper reports the process that led to the implementation of a Business Intelligence tool for a small commercial company that basically operates as a trading company, intermediary between manufacturing companies and consumer or distributor companies, in import operations of chemicals and plastic resins. The company has the perception that the implementation of the tool increased its volume of business. . It was confirmed the importance of using a framework to choose the solution that is appropriate to the needs of the company and the difficulties of small organizations to deploy software tools without the support of external personnel. The work aims to share the gained experience, so that it can be useful to researchers and professionals involved with the theme.

KEY WORDS

Business Intelligence; small business; implantation.

INTRODUÇÃO

Cada vez mais a Tecnologia da Informação (TI) vem assumindo um papel estratégico nas organizações, tornando imperativo que as empresas as utilizem para operar de forma eficiente e eficaz, permanecendo competitivas.

Laudon e Laudon (2015) corroboram essa informação ao dizer que das ferramentas que os administradores dispõem, as tecnologias e os sistemas de informação estão entre as mais importantes para atingir altos níveis de eficiência e produtividade nas operações, especialmente quando combinadas com mudanças no comportamento da administração e nas práticas e processos de negócio.

As tecnologias de Sistemas de Informação já são utilizadas há muito tempo por grandes empresas, porém nota-se que a realidade de empresas de pequeno porte é diferente; para a maioria destas empresas, é relativamente recente a implantação de tecnologias e sistemas de informação, o que pode ser justificado pelo fato de que até pouco tempo havia a ausência do entendimento, por parte dos administradores dessas empresas, sobre a importância desses sistemas para a melhoria de seus processos, além dos elevados custos de implantação dos mesmos, não só em termos de *hardware* e *software*, como também de serviços a eles relacionados.

O objeto deste estudo é uma empresa comercial de pequeno porte que opera basicamente como uma *trading company*, intermediária entre empresas fabricantes e empresas consumidoras ou distribuidoras, em operações de importação de matérias primas e insumos, principalmente produtos químicos e resinas plásticas; por questões de sigilo, a empresa é chamada, neste trabalho, de XYZ, localizando-se em cidade próxima à capital do Estado de São Paulo; existe há dez anos e conta com nove funcionários. O proprietário da empresa tinha a percepção de que oportunidades de negócio estavam sendo perdidas face à ausência de informações estruturadas sobre interesses dos clientes, sendo essa a principal motivação do projeto aqui descrito.

OBJETIVO E ASPECTOS METODOLÓGICOS

Dado esse cenário, elaborou-se este trabalho, que teve como objetivo relatar o processo que levou à implementação de uma ferramenta de *Business Intelligence* para a empresa já apresentada acima, de forma a que a experiência adquirida possa ser útil a pesquisadores e profissionais envolvidos com o tema.

O trabalho, do ponto de vista metodológico, pode ser classificado como uma pesquisa-ação, nos termos propostos por Grundy e Kemmis (1982): “identificação de estratégias de ação planejada que são implementadas e, a seguir, sistematicamente submetidas a observação, reflexão e mudança” – o autor participou desse processo na condição de consultor independente, contratado pela XYZ para confirmar a percepção de seu proprietário acerca dos problemas vividos pela mesma, propor soluções para os mesmos e implementá-las.

O DESENVOLVER DO TRABALHO – O PROCESSO DE SELEÇÃO

O trabalho iniciou-se com a realização de entrevistas não estruturadas com o proprietário da empresa e seu imediato; havia o sentimento de que oportunidades de negócio estavam sendo perdidas face à ausência de informações estruturadas sobre produtos que interessavam aos clientes. Conforme Marconi e Lakatos (2015) que dizem que esse tipo de entrevista é útil para explorar mais amplamente uma questão, com perguntas abertas que podem ser respondidas dentro de uma conversação informal, podendo cada situação abordada ser desenvolvida em qualquer direção que se considere adequada.

Em paralelo, desenvolveu-se um trabalho de observação assistemática, envolvendo os personagens acima citados e mais um funcionário da XYZ. Os mesmos autores (MARCONI & LAKATOS, 2015) definem essa técnica aquela em que se recolhe os fatos sem que sejam necessários meios técnicos especiais para registro do observado; dizem também que o êxito da aplicação dessa técnica deve-se ao preparo e à experiência, do observador – essas fases do trabalho (entrevistas e observação) consumiram cerca de 8 semanas.

Nesse caso, a longa experiência do consultor, com vivência na área e em projetos de todos os portes, permitiu considerar que essas técnicas de pesquisa seriam adequadas para a aquisição dos conhecimentos necessários ao prosseguimento do projeto.

O que se constatou, como principal problema da empresa, foi que TI era utilizada apenas nos processos de natureza administrativa, não sendo empregada para a prospecção de novos clientes e negócios, com a empresa operando basicamente de forma reativa, cotando produtos quando eventuais compradores solicitavam e registrando apenas as vendas efetuadas, deixando de registrar os pedidos de cotação.

O proprietário da XYZ tinha a percepção de que a implementação de uma ferramenta ERP (Enterprise Resource Planning) poderia resolver o problema acima citado e preparar

a empresa para o crescimento, sendo esse um aplicativo que atenderia a XYZ de forma global. De acordo com O'Brien e Marakas (2013) e Moller (2005), os ERP podem ser definidos como sistemas de informações adquiridos na forma de pacotes comerciais de *software* que permitem a integração dos sistemas de informações transacionais e dos processos de negócios de uma dada organização, objetivando conferir à mesma vantagens competitivas, permitindo-lhe reagir de forma rápida e flexível aos acontecimentos externos; são constituídos por diversos módulos que atendem a funções empresariais específicas e baseados em bancos de dados integrados.

Seguindo a estratégia padrão proposta por Breternitz e Galhardi (2011), foi elaborada uma RFI (*Request for Information*), enviada a potenciais fornecedores, buscando informações acerca dos pontos mais importantes a serem considerados na seleção de uma ferramenta ERP, como por exemplo, funcionalidades, ambiente operacional, custos, casos de implantação com sucesso etc.

Essa RFI foi construída com base nas especificações dos requisitos necessários a uma ferramenta que atendesse às necessidades da XYZ, conforme especificado por Sommerville (2016): os requisitos podem ser definidos como as descrições do que o *software* deve fazer, os serviços que ele deve oferecer e as restrições ao seu funcionamento.

O passo seguinte foi a análise das respostas recebidas, tendo o item “custos” inviabilizado a aquisição desse tipo de solução. Aqui, entende-se custos de forma abrangente, não apenas as licenças de *software*, mas também necessidades de consultoria para implementação, *hardware*, treinamento etc.

Resolveu-se então buscar ferramenta de Business Intelligence, a ser empregada basicamente para a prospecção de novos clientes e negócios, permitindo que a empresa deixasse de operar de forma reativa, apenas cotando produtos para eventuais compradores e registrando as vendas efetuadas, o que era, na percepção do empresário, sua prioridade maior. Davenport (2006, p. 106) define *Business Intelligence* como

The term “business intelligence,” which first popped up in the late 1980s, encompasses a wide array of processes and software used to collect, analyze, and disseminate data, all in the interests of better decision making. Business intelligence tools allow employees to extract, transform, and load (or ETL, as people in the industry would say) data for analysis and then make those analyses available in reports, alerts, and scorecards

Ainda seguindo a estratégia acima mencionada, passou-se a buscar no mercado a melhor ferramenta (*best-of-breed*) para a solução do principal problema da empresa, à qual poderiam ser conectadas, na medida das necessidades, ferramentas que atendessem a outras necessidades da XYZ, tentando-se dar a elas um certo grau de integração. Note-se que apesar das prováveis dificuldades nessa conexão e integração, as soluções *best-of-breed* quase sempre possuem algumas vantagens em relação aos ERP, dentre elas melhores funcionalidades, mais flexibilidade, não dependência a um único fornecedor etc. (STEFANO, 2000). Esse ponto de vista é endossado por Panorama (2017, p. 1):

Best of breed options used to have negative connotations. Nowadays, however, integration tools have evolved to the point that many of the historic risks of best of breed solutions aren't as big of a threat as they used to be. At the same time, single ERP systems aren't always the best fit for organizations. Some organizations will find that single ERP system option makes the most sense for them, while others will find that best of breed is the best fit. It is important to weight the risks, tradeoffs, and pros and cons of each to determine the best solution for you.

Essa busca também seguiu a estratégia padrão proposta por Breternitz e Galhardi (2011): foi elaborada uma RFI, enviada a potenciais fornecedores, buscando informações acerca dos pontos mais importantes a serem considerados na seleção de uma ferramenta de *software*; essa segunda RFI focou-se especialmente no que seria necessário para atender às necessidades de prospecção de novos clientes e negócios, podendo a empresa passar a atuar de forma proativa, oferecendo seus serviços ao mercado ao invés de simplesmente atender a pedidos de cotações e eventualmente fechar negócios oriundos desses pedidos. Dando sequência ao trabalho, foram analisadas as respostas dadas por possíveis fornecedores; àqueles cujos produtos pareciam atender às necessidades da XYZ, foi enviada uma RFP (*Request For Proposal*), um convite a esses potenciais fornecedores para que submetessem à XYZ uma proposta para implantação de seu sistema. Nessa etapa, foram detalhadas e complementadas as informações obtidas na etapa anterior, considerando-se adicionalmente aspectos como detalhamento de custos, prazos de implantação, necessidades de customização e outros.

Para a análise das propostas recebidas, foi usado o *framework* proposto por Breternitz e Galhardi (2011), com ênfase nos aspectos que se seguem:

Funcionalidades e aderência

Funcionalidade é quase sempre o enfoque principal da maioria das avaliações, mas a prática recomenda que esse critério não seja supervalorizado, pois quando isso acontece outros fatores importantes acabam sendo deixados de lado. O objetivo maior da análise deste ponto é avaliar o grau de aderência e adequação da ferramenta aos processos de negócio da empresa e ao seu modelo de gestão; como a ferramenta não iria a princípio substituir ou interagir com outros aplicativos e se pretendia atender a necessidades claramente explicitadas nessa altura do processo, não foram encontradas dificuldades nesse aspecto.

Custos

Especialmente no caso das empresas de pequeno porte, expectativas realistas acerca do cronograma de desembolsos, do ROI (*Return on Investment*) e do TCO (*Total Cost of Ownership*) são essenciais para decisão quanto ao prosseguimento do projeto e posterior avaliação dos resultados da implantação.

Serviço e Suporte

As pequenas empresas são hoje um alvo prioritário para maioria dos fornecedores de aplicativos (mesmo os grandes fornecedores), o que está fazendo com que estes reduzam substancialmente seus preços para poder atingir este segmento. Corrêa (1998), no entanto, lembra que vale um raciocínio simples: para ganhar o mesmo dinheiro vendendo soluções mais baratas, é necessário que um número muito maior clientes seja atendido. Isso pode significar um grau de atenção bem menor para cada um dos novos clientes do que aquele que esses fornecedores dedicavam a clientes de porte maior, numa época em que as soluções podiam ser comercializadas por um preço unitário muito maior e, portanto, não eram necessários tantos clientes, mas apenas alguns de grande porte. No caso da XYZ, pretendia-se adquirir solução fornecida por empresa de porte tal que as necessidades de serviço e suporte pudessem ser atendidas de forma plena.

Estabilidade do fornecedor

Tendo em vista a importância que a solução escolhida teria para a XYZ, a estabilidade de um possível fornecedor era algo a ser verificado de forma acurada. Imagine-se os custos e o trauma de uma nova seleção e implantação de um aplicativo se, pouco tempo após a implantação de um sistema, o fornecedor simplesmente se retirar do mercado ou descontinuar seu produto,

Usabilidade

Segundo Medeiros Jr *et al.* (2009), quando um sistema é amigável (fácil de ser utilizado), torna-se mais fácil treinar e manter treinados seus usuários, o que torna sua implantação mais fácil e manutenção, no médio e longo prazos, mais barata, por exigir menos retrabalho e esforços de treinamento. São desejáveis recursos amigáveis, como interfaces gráficas, menus inteligentes etc.

Implantação

A implantação não é propriamente um critério para escolha de uma solução como a objeto deste texto, mas a postura dos possíveis fornecedores acerca do assunto deve ser considerada. Este é um ponto que frequentemente só começa a ser discutido após a escolha da ferramenta, o que pode gerar surpresas muito desagradáveis: frequentemente os custos de implantação não são analisados durante o processo de seleção do sistema. As empresas, especialmente as pequenas quase sempre não conseguem implantar aplicativos sem o apoio de consultorias e/ou do fornecedor. Cronogramas de implantação, com clara definição das responsabilidades de todos devem ser elaborados, custos e necessidades de treinamento avaliados, etc., mas tudo ainda antes da escolha final do fornecedor, como já se disse.

A análise das propostas, complementada por discussões com os possíveis fornecedores, levou à escolha da ferramenta GoodData TOTVS Smart Analytics (GD), que é vendida pelo fornecedor como uma solução de Business Intelligence, que pode vir a ser concatenada a outras ferramentas, aumentando seu escopo.

A ferramenta GD, que conta com cerca de trinta mil usuários em todo o mundo, é comercializada na modalidade SaaS (Software as a Service – o sistema reside na nuvem

e não no *hardware* da XYZ) e foi desenvolvida pela GoodData, empresa baseada no Vale do Silício e representada no Brasil pela Totvs, que desenvolveu interface para ligá-la a seu ERP Proteus.

No caso da XYZ, os dados serão entrados manualmente, face às características operacionais da mesma e as informações produzidas a partir dos dados poderão ser acessadas via *dashboards* (indicadores ou gráficos que transmitem informações visualmente) ou consultas através de telas ou relatórios impressos produzidos pelos usuários do sistema – o sistema permite que os usuários especifiquem essas telas e relatórios e depois, facilmente, os gerem.

Além disso, se no futuro ocorrer a necessidade de implementação de um ERP, o GoodData pode ser conectado ao mesmo de forma simples, desde que, evidentemente, seja escolhida a ferramenta Proteus.

O PROCESSO DE IMPLANTAÇÃO

Para se obter o melhor rendimento do aplicativo, foi necessário definir e implementar processos estruturados de trabalho e definir responsabilidades por atividades como registrar os contatos iniciais efetuados com possíveis clientes e todas as interações realizadas com estes, em especial pedidos de cotação, negócios fechados, razões da perda de negócios (indisponibilidade de produtos, custos etc.), rotinas de consultas a serem efetuadas no dia-a-dia etc.

A implementação foi conduzida com apoio do pessoal da Totvs, que atuou inclusive na definição dos processos internos e na especificação das saídas do sistema (*dashboards*, telas e relatórios) e no treinamento dos usuários; foram utilizadas 120 horas de trabalho de pessoal da Totvs nesse processo, mais as horas de trabalho do consultor.

Da análise das RFI/RFP, ficou claro que a estrutura tecnológica de que dispunha a empresa precisava ser atualizada; assim trabalho compreendeu também a especificação de recursos de *hardware*, *software*, telecomunicações (acesso à internet), suporte e *back-up* necessários à operação da ferramenta, pois os recursos anteriores eram inadequados.

Para o atendimento dessas necessidades foram solicitados orçamentos para três empresas situadas na cidade onde se localiza a XYZ, tendo se optado por ambiente Windows em função da facilidade de suporte e operação nesse ambiente, com o qual já estava familiarizado o pessoal da empresa.

Em termos de *hardware* especificou-se, sem maiores estudos, equipamentos de fornecedores tradicionais (Acer, Asus, Dell, HP e Lenovo), por ser do senso comum serem esses equipamentos similaridades em termos de performance e qualidade, tendo a decisão sido tomada de acordo com critérios de custo.

Quanto a suporte (instalação e manutenção de *hardware* e *software* e apoio aos usuários), optou-se pela manutenção da empresa que já prestava serviços à XYZ, que se encontrava satisfeita com a qualidade e os custos praticados por esse fornecedor.

Para fins de *back-up* optou-se pela utilização de serviços de armazenagem na nuvem (*cloud*) – situação em que todos os dias, ao final do expediente, os dados da empresa são transmitidos, via Internet, aos computadores do prestador desse serviço, podendo os mesmos serem recuperados em caso de sinistro ou de interesse da XYZ em permitir que os dados sejam acessados de fora do ambiente da empresa. O Google Drive foi o serviço escolhido, em função da familiaridade do pessoal da XYZ com o produto e por não serem seus custos relevantes. A alternativa era a utilização de fitas magnéticas, que exigiam maior atenção e cuidados, como por exemplo a retirada dos cartuchos do ambiente da empresa, inclusive tornariam menor a segurança dos dados da XYZ.

CONCLUSÕES

O proprietário da XYZ tem a percepção de que a implementação da ferramenta incrementou seu volume de negócios, embora não consiga mensurar exatamente a relação entre esse incremento e o uso da ferramenta.

Confirmou-se a importância da utilização de um *framework* para a escolha de solução adequada às necessidades da empresa usuária – o consultor atua nesse momento em outro projeto onde uma escolha de solução *quick pick* feita anteriormente está exigindo da empresa novos investimentos para suportar uma mudança no patamar de operação, simplesmente precisando abandonar investimentos em *software* feitos anteriormente. *Quick pick* é a escolha de uma solução feita a partir de uma análise rápida, superficial, da situação da empresa e das soluções disponíveis no mercado (BRETERNITZ & GALHARDI, 2011).

Os mesmos autores ressaltam as dificuldades de as organizações implantarem ferramentas de *software* sem o apoio de pessoal externo; também no caso da XYZ essa participação foi fundamental, tendo sido as responsabilidades, prazos e custos claramente estabelecidos em contratos assinados entre a empresa e a Totvs.

REFERÊNCIAS BIBLIOGRÁFICAS

- BRETERNITZ, V. J. & GALHARDI, A. C. Contribuições ao processo de seleção de sistemas ERP (Enterprise Resource Planning) para pequenas e médias empresas. **Revista Eletrônica de Tecnologia e Cultura**. 3(2). 2011.
- CORRÊA, H. L. Aspectos a se considerar na seleção e implantação de uma solução ERP para médias empresas. **Guideline Gestão Empresarial Computerworld**, edição de 27 de julho de 1998.
- DAVENPORT, T.H. Competing on Analytics. **Harvard Business Review**, 84(1). 2006.
- GRUNDY, S. J. Three modes of action research. **Curriculum Perspectives - Geelong**, 2(3). 1982.
- LAUDON, K. C. & LAUDON, J. P. **Sistemas de informação gerenciais**. 11ª ed. São Paulo: Pearson Prentice Hall, 2015.
- LAKATOS, E. M. & MARCONI, M. A. Técnicas de pesquisa: planejamento e execução de pesquisas, amostragens e técnicas de pesquisa, elaboração, análise e interpretação de dados. 7ª ed. São Paulo: Atlas. 2015.
- MEDEIROS JR. A. et al. **Proposta de um modelo para a seleção de ERPs**. Disponível em www.ead.fea.usp.br/semead/12semead/resultado/trabalhosPDF/607.pdf. Acesso em 15.06.2010. 2009.
- MOLLER, C. ERP II: a conceptual framework for next-generation enterprise systems. **Journal of Enterprise Information Management**, 18(4). 2005.
- O'BRIEN, J. A. & MARAKAS, G. M. **Administração de sistemas de informação**. 15ª ed. São Paulo: McGraw-Hill, Bookman. 2013.
- PANORAMA CONSULTING. Five variables to create a digital transformation strategy that best fits your organization. Greenwich Village: Panorama Consulting. 2017.
- SOMMERVILLE, I. **Software engineering**. 10ª ed. Harlow: Pearson. 2016.
- STEFANOU, C. J. The Selection Process of Enterprise Resource Planning (ERP) Systems. In: AMCIS 2000, 418. Long Beach. **Proceedings**. 2000.

ANÁLISE DE UM ALGORITMO DESENVOLVIDO PARA SOLUÇÃO DE DEEP LEARNING UTILIZANDO REDES NEURAS CONVOLUCIONAIS PARA DIFERENCIAR GATOS DE CACHORROS

ANALYSIS OF A DEVELOPED ALGORITHM FOR DEEP LEARNING SOLUTION USING CONVOLUTIONAL NEURAL NETWORKS TO DIFFERENTIATE CATS

Jonas de Souza TEGA (Aluno)

jonas_tega@hotmail.com

Ciências da Computação, Centro Universitário Padre Anchieta

Carlos Eduardo CÂMARA (orientador)

ccamara@anchieta.br

dinhocamara@gmail.com

Ciências da Computação, Centro Universitário Padre Anchieta

Resumo

Com o passar dos anos, a inteligência artificial passou por diversas evoluções, se segmentado em diversas áreas. Um desse ramos é dedicado a aprendizagem de máquina, que por sua vez vem se destacando com o aprendizado profundo (*Deep Learning*), principalmente na parte de reconhecimento de imagens. Uma das estratégias utilizadas para atingir esse destaque, é utilizar as redes neurais convolucionais para que a máquina concretize seu aprendizado. Sendo assim, este artigo trará os conceitos básicos sobre redes neurais convolucionais, e como a sua fusão com aprendizado profundo se encaixa tão bem para reconhecimento de imagens.

Palavras-Chave

Inteligência Artificial; Aprendizado Profundo; Reconhecimento; Redes Neurais Convolucionais.

Abstract

Over the years, artificial intelligence has undergone several evolutions, if segmented into several areas. One of these branches is dedicated to machine learning, which in turn has been highlighted by Deep Learning, mainly in the image recognition part. One of the strategies used to achieve this highlight is to use convolutional neural networks to make the machine realize its learning. Therefore, this article will introduce the basics of convolutional neural networks, and how their fusion with deep learning fits so well for image recognition.

Keywords

Artificial Intelligence; Deep Learning; Recognition; Convolutional Neural Networks.

INTRODUÇÃO

Desde de muito tempo atrás, o ser humano tenta compreender como funciona a inteligência humana e com o passar dos tempos esse conhecimento foi se formalizando, dando início a uma nova investigação, que visa avançar os conhecimentos adquiridos sobre inteligência, para formar algo que pensa desde a sua essência, criando assim o campo da Inteligência Artificial.

Com o passar dos anos, e o avanço da tecnologia, o aprendizado de máquina passou a ser cada vez mais comum, principalmente após a segunda guerra mundial, e hoje em dia é um dos assuntos mais pesquisados e desenvolvidos no ramo da tecnologia da informação. Assim, inúmeras pesquisas foram desenvolvidas para máquinas efetuar determinadas ações, ao ponto em que as máquinas possuem a capacidade de desenvolver determinado conhecimento sobre uma ação, podendo desenvolver suas próprias técnicas para melhor efetuar a ação.

Atualmente o estudo de inteligência artificial se divide em diversas áreas, focando em diferentes formas de aprendizado de máquina. Dentre eles um que vem se destacando é o *Deep Learning*, que tem como objetivo uma aprendizagem mais profunda que, através de um conjunto de algoritmos de Redes Neurais Artificiais, modelam melhor uma quantidade bem superior de dados, visando o aprendizado de máquinas através da observação dessa massa de dados. Isso permite, por exemplo, aprofundar seus níveis de comparação, a ponto de conseguir diferenciar certas imagens, podendo assim trazer um resultado de busca muito mais preciso e rápido.

Esse tipo de técnica é utilizado por plataformas de busca, como o Google utiliza para realizar busca de imagens, até para plataformas de segurança que utiliza reconhecimento facial de pessoas em meio à multidão, ou para um sistema de segurança.

Através dessa crescente pesquisa em torno desse tema, esse projeto analisa os componentes básicos para que uma Rede Neural Densa (*Deep Learning*) consiga operar, e avaliar os resultados que ela permite no reconhecimento automático de imagens.

A maneira utilizada para fazer essa análise, foi emprestar um código feito por Kinsley, disponível em *How to use your trained model: Deep Learning basics with Python, TensorFlow and Keras p.6* [16], que diferencia se uma determinada imagem é um cachorro, ou um gato, explicando detalhadamente, e testando mais de um modelo de Redes Neurais Densa, a fim de perceber como elas operam, e como suas estruturas funcionam. Todas as estruturas utilizadas no código, serão descritas, para maior entendimento do mesmo.

CONCEITOS BÁSICOS

INTELIGÊNCIA ARTIFICIAL

“A inteligência artificial (IA) é simplesmente uma maneira de fazer o computador pensar inteligentemente. Isso é conseguido estudando como as pessoas pensam quando estão tentando tomar decisões e resolver problemas, dividindo esses processos de pensamento em etapas básicas e desenhando um programa de computador que solucione problemas

usando essas mesmas etapas. A IA então fornece um método simples e estruturado de se projetar programas complexos de tomada de decisão”. (LEVINE, DRANG, EDELSON 1988)

Seu início se deu logo no fim da segunda guerra, em 1950 quando Alan Turing fez uma proposta chamada de teste de Turing, na qual se propõe avaliar se uma máquina que possui uma inteligência satisfatória ou não. Para fazer a qualificação, o teste envolve uma conversa entre duas pessoas em uma máquina, porém cada um está isolado do outro. Uma das pessoas faz perguntas, como em um diálogo normal, e dessas perguntas veem respostas simultâneas, tanto da máquina, quanto da outra pessoa. Caso o questionador não conseguisse diferenciar com certeza qual a resposta estava vindo da máquina, e qual vinha da outra pessoa, a mesma era classificada com uma inteligência satisfatória, já que seu sistema de pensamentos para responder as perguntas não conseguia ser diferenciado do humano. [1,9]

Para a máquina conseguir êxito no teste, ela precisaria possuir processamento de linguagem natural, representação de conhecimento, raciocínio automatizado, aprendizado de máquina. Possuindo essas quatro características, a máquina seria capaz de se comunicar, armazenar informações, comparar e manipular uma conclusão e por fim aprender tudo aquilo que está sendo absorvido nas outras três características. Essa deveria a única forma de uma máquina passar no teste. [1]

Um outro teste proposto, o Teste de Turing Total, se baseava na mesma ideia, porém dessa vez o questionador iria interagir diretamente com a máquina. O que diferenciaria as circunstâncias do outro teste, assim para conseguir executar tal alteração, a máquina precisaria ter uma tela, com possibilidade de o computador receber objetos oferecidos pelo interrogador, essa tese é chamado de teste de Turing total. [1]

A partir deste teste, surgem mais dois conceitos essenciais para a IA sendo eles a visão computacional, para receber objetos, e a robótica, para manipular e movimentar-se, assim passaria a estudar a forma como um computador “observa”, “percebe” e manipula objetos, junto com a robótica, trazendo o estudo de que uma máquina de forma autônoma possa manipular objetos precisamente e, na maioria dos casos, fazendo-o de uma forma que supera a capacidade humana. [1]

REDES NEURAIIS ARTIFICIAIS

Começou a ser fortemente pesquisada nos anos 90, apesar de já se ter havido algumas publicações na década de 50. As redes neurais possuem o intuito de resolver diversos tipos de problemas, porém obtendo uma alta capacidade de mapear sistemas não lineares, e através dele aprender os comportamentos envolvidos através dos dados obtidos. [4]

“Redes neurais artificiais são modelos computacionais inspirados no sistema nervoso de seres vivos. Possuem a capacidade de aquisição e manutenção do conhecimento (baseado em informações e podem ser definidas como um conjunto de unidades de processamento, caracterizadas por neurônios artificiais, que são interligados por um grande número de interconexões (sinapses artificiais), sendo representadas aqui por vetores/matrizes de pesos sinápticos”. (SILVA, 2016)

A primeira publicação foi em 1943, escrita por *McCulloch & Pitts*, analisando um neurônio biológico e reproduzindo um modelo matemático inspirado nesse neurônio biológico. Assim surgiu a primeira ideia de um neurônio artificial. *Em 1949, Hebb propôs o primeiro método de treinamento para uma rede, baseando-se em hipóteses e observações de caráter neurofisiológico.* Outros grandes marcos dessa época se deram no fim da década de 50 e início da década de 60, com o desenvolvimento do primeiro neuro-computador denominado *Mark I – Perceptron*. Outro modelo de redes neurais é a

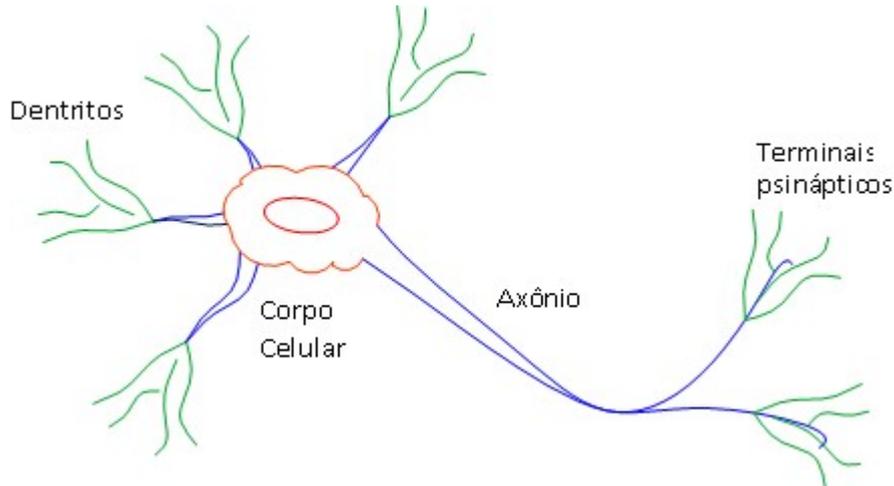
rede *Adaline* (*ADaptative Linear Element*) e, posteriormente, *Madaline*, podendo pôr em prática as ideias de redes neurais artificiais e deixando essas ideias cada vez mais palpáveis. Porém a publicação de *Perceptrons - an introduction to computational geometry* em 1969 por *Minsky & Papert*, “*acabou colocando os estudos da área de neurocomputação em um grande inverno*”. *Nesta publicação foi exposta uma limitação das redes neurais artificiais, por serem constituídas de apenas uma única camada, e por não conseguirem aprender um relacionamento entre as entradas e saídas de funções lógicas bem simples como XOR (ou exclusivo), por exemplo. Com essa demonstração mostrava-se a impossibilidade dessas redes realizarem a correta classificação de padrões para classe não linearmente separáveis.* Passaram os anos, computadores evoluíram, algoritmos foram otimizados, e os estudos sobre o sistema nervoso biológico tiveram novas descobertas. *Com esses fatores Rumelhart, Hinton e Williams, publicam Parallel Distributed Processing, no qual eles desenvolveram um algoritmo intitulado backpropagation, que permitia ajustar os pesos em uma rede com mais de uma camada, solucionando os problemas referentes ao XOR, e quebrava assim a barreira imposta em 1969. A partir de então, muitos estudos passaram a se desenvolver na área novamente.* [4]

As principais características de uma rede neural artificial, adaptação por experiência, capacidade de aprendizado, habilidade de generalização, organização de dados, tolerância a falhas, armazenamento distribuído e facilidade de prototipagem. Atualmente pode-se aplicar os conceitos de redes neurais em diversas áreas, porém a que mais vem se destacando acaba sendo a área de reconhecimento de padrões (utilizada para análise de determinadas ações, ou utilizada na área de segurança para reconhecimento facial, ou da saúde para reconhecimento de diagnósticos de cânceres e várias outras doenças automaticamente), e de agrupamento de dados (muito utilizado para agrupar diversos dados espalhados pela internet, e filtrá-los, mostrando somente dados válidos para a análise). [4, 5, 11]

NEURÔNIO BIOLÓGICO

É uma célula nervosa, Figura 1, pode ser definida como o processador biológico, altamente estimulável, age em paralelo através de impulsos eletroquímicos, com a funcionalidade de processar, transmitir, pensar, e memorizar os dados e informações adquiridas através dos nossos receptores (olhos, mãos, boca, etc.). [4,6]

Figura 1. Fisiologia de um neurônio biológico. [6]

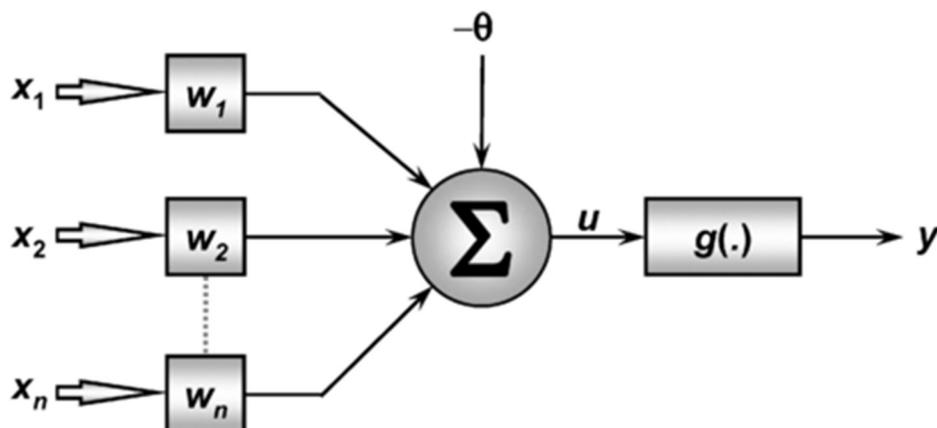


A responsabilidade de processar todas as informações advindas dos dendritos é do corpo celular. Os dendritos, são ramificações localizadas nas pontas dos neurônios. Após processada a informação é transmitida através do axônio por impulsos elétricos, que por final repassa a informação necessária para os próximos neurônios através de suas terminações sinápticas, Figura 1. [4,6]

NEURÔNIO ARTIFICIAL

Inspirado no neurônio biológico, o neurônio artificial, Figura 2, tenta emular a mesma tarefa da sua inspiração, assim, basicamente, ele receberá a informação, irá processá-la, em seguida irá repassá-la processada para outro neurônio. As interligações de vários neurônios formam uma rede denominada rede neural, com paralelismo e alta conectividade. [4]

Figura 2. Fisiologia de um neurônio artificial. [4]



Sua funcionalidade se baseia nos sinais de entrada $\{x_1, x_2, x_n\}$ sendo multiplicadas pelos pesos sinápticos na entrada do neurônio, $\{w_1, w_2, \dots, w_n\}$, onde está especificado se a informação apresentada na entrada é relevante ou não. Em seguida, todos os resultados adquiridos através destes pesos são somados no combinador aditivo linear $\{\Sigma\}$, e, após obter a soma, ela passa pela soma de um limiar de ativação $\{-\theta\}$, sendo o resultado gerado, o valor do potencial de ativação $\{u\}$, caso $u > -\theta$, o neurônio produzirá um potencial excitador, caso contrário, será inibidor; Por fim, u passa por uma função de

ativação $\{g\}$, limitando o valor adquirido no começo do processo, gerando o sinal de saída $\{y\}$, que poderá ser o resultado final de uma rede, ou o sinal de entrada de um próximo neurônio. [4,7]

As expressões do resultado gerado por um neurônio, propostas por Pitts e McCulloch são:

$$u = \sum_{i=1}^n w_i \cdot x_i - \theta$$

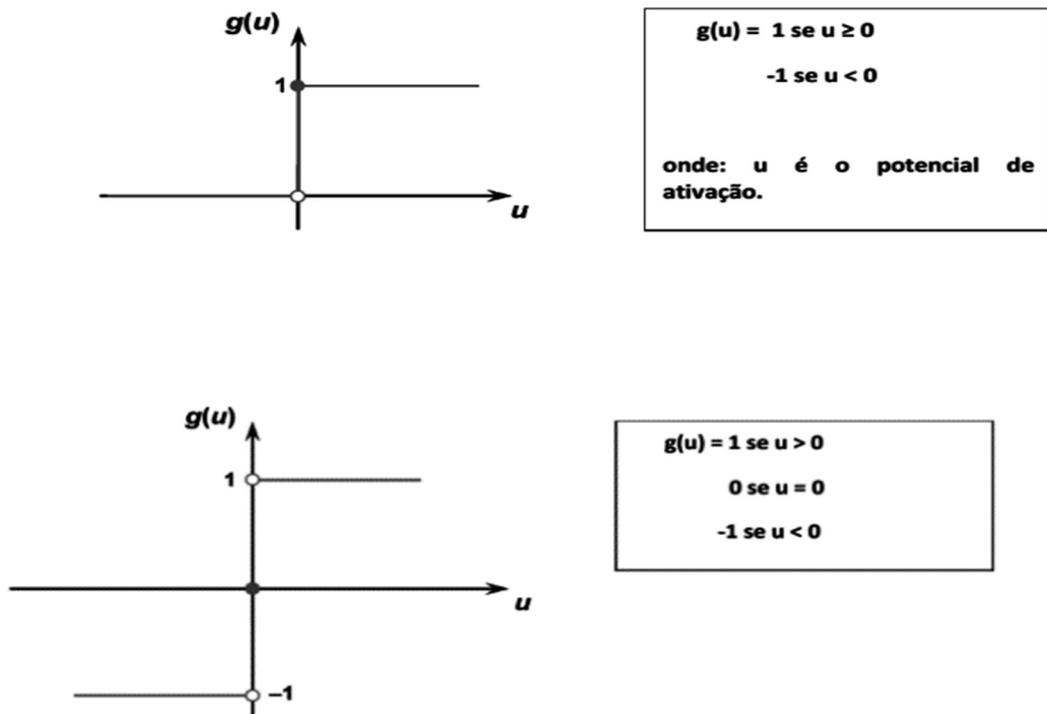
$$y = g(u)$$

“As funções de ativação podem ser divididas em dois grupos, sendo parcialmente diferenciáveis, e totalmente diferenciáveis.” [4]

Quando as derivadas de primeira ordem forem inexistentes, é considerada parcialmente diferenciável, já que as diferenças menores que o ponto a serem pesquisados são inexistentes. Esta classe possui três funções principais: função degrau, função degrau bipolar, função rampa simétrica. [4]

Nas funções degrau, e degrau bipolar, Figura 3, são basicamente iguais; ambas analisam o resultado, porém no degrau, o neurônio só é ativado quando possui um valor maior que 0, enquanto a degrau bipolar possui o valor nulo do neurônio, que será quando ele atingir o valor 0. [4, 13]

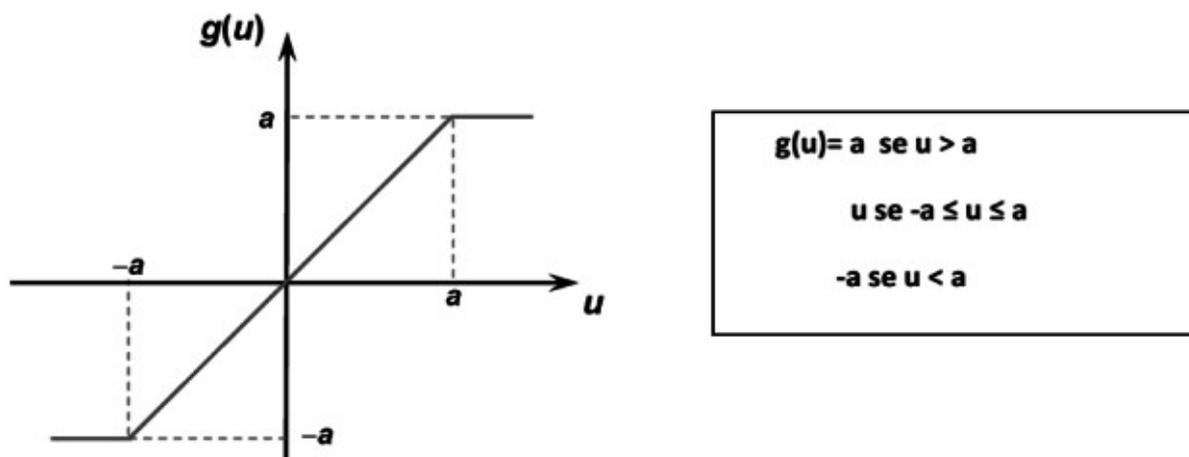
Figura 3. Representação gráfica das funções degrau, e degrau bipolar. [13]



Já a função rampa simétrica, Figura 4, possui um diferencial das outras duas, já que essa ao invés de fazer uma análise entre 0 e 1, ela determina qual o espaço será o de ativação,

assim essa função mostra a evolução da ativação desde seus valores negativos, até seus valores positivos, criando assim a rampa. [4, 13]

Figura 4. Representação gráfica da função rampa simétrica. [13]



Para a função ser totalmente diferenciável, as derivadas de primeira ordem são conhecidas e analisadas em todo domínio da função. Possui quatro funções mais conhecidas dessa divisão.

A primeira e a segunda são muito parecidas, sendo elas função logística, e tangente hiperbólica, Figura 5. A diferença básica entre elas é que a função logística, $0 < g(u) < 1$, sendo que $g(u)$ tende a 1, ou seja, os valores 0 e 1 são assíntotas. A função tangente hiperbólica, $-1 \leq g(u) \leq 1$. O mesmo para esta função, ou seja, -1 e 1 a curva é assintótica a -1 e a 1. Elas utilizam uma expressão, e seus valores são analisados entre 0 e 1 na primeira, e na segunda entre -1 e 1. [4, 13]

Outra função utilizada nessa divisão, é a gaussiana, Figura 6, que consiste em analisar a saída do neurônio, fazendo com que ele seja igual aos valores que possui o mesmo potencial de ativação, $\{u\}$, e que esteja na mesma distância de seu centro. Para marcar o centro da função, utiliza-se o valor de c (a média), e o desvio padrão, para analisar se ambas possuem o mesmo valor. [4, 13].

Figura 5. Representação gráfica da função logística. [13]

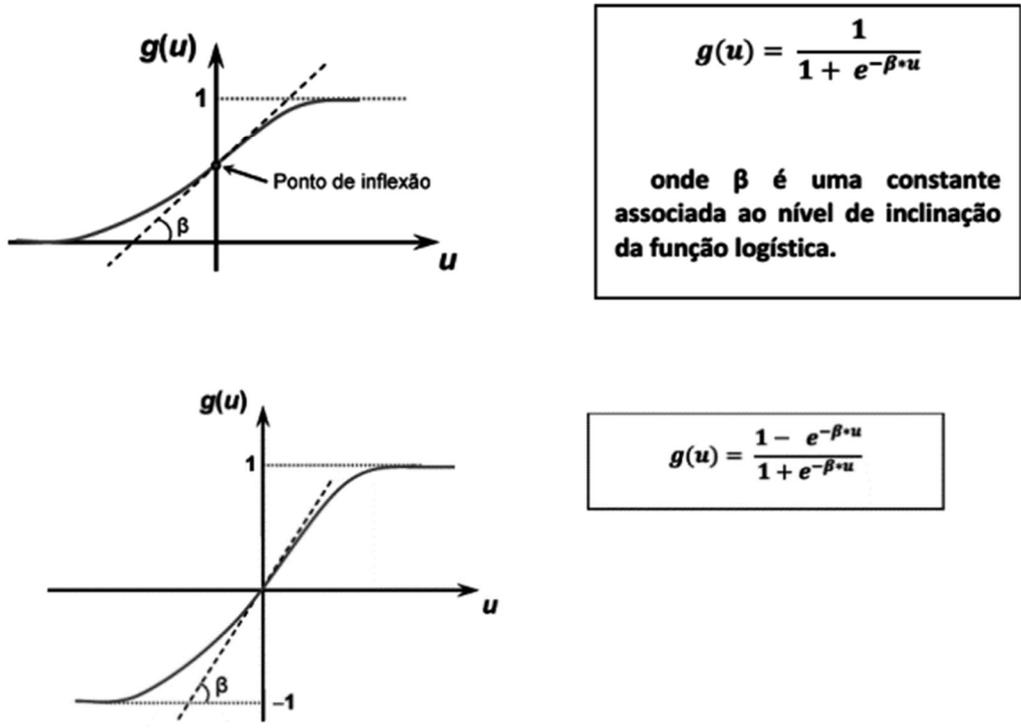
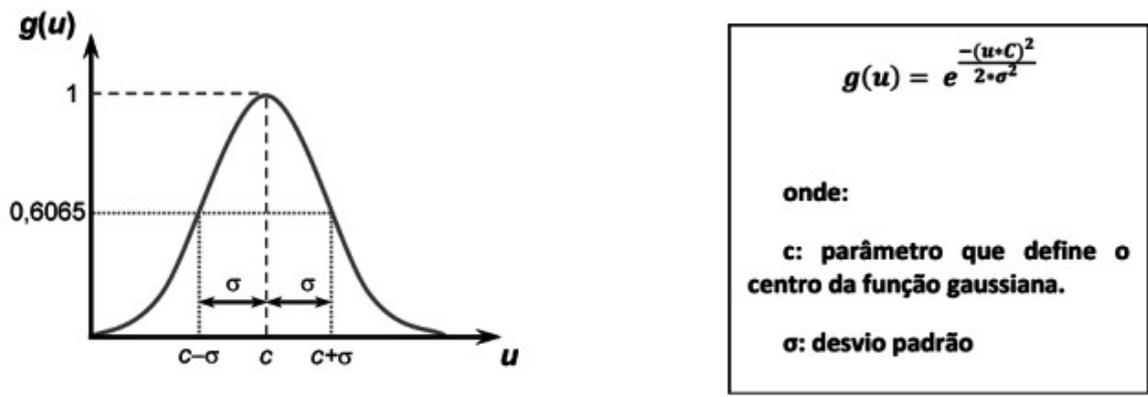
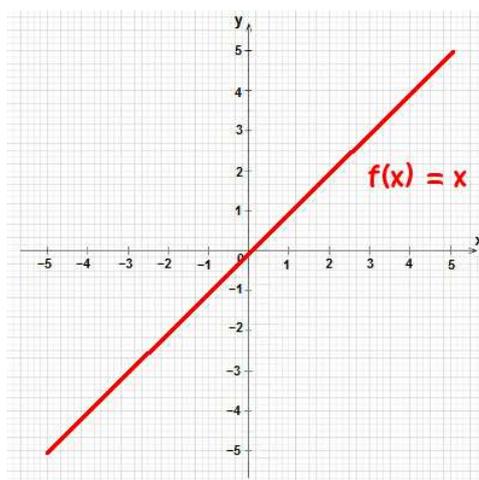


Figura 6. Representação gráfica da função gaussiana. [13]



Por fim, a última e a função linear, que acaba sendo utilizada para mapear o comportamento entre as variáveis de entrada e saída, já seus valores são iguais ao potencial de ativação. [4, 15]

Figura 7. Representação gráfica da função linear. [15]



DEEP LEARNING (APRENDIZADO PROFUNDO)

Baseado nos princípios do teste de Turing, o aprendizado profundo passa a ser uma evolução do aprendizado de máquina, logo, seu intuito é criar mecanismos que diante de um problema possam resolvê-lo a partir de uma grande análise previamente já feita pela máquina; podendo assim em um caso de reconhecimento de imagem fazer com que a máquina reconheça o que foi pré-determinado, através de um mecanismo treinado por ela, trazendo assim a resposta mais próxima do esperado. [10]

De acordo com Deng e Yu, “o aprendizado profundo são várias técnicas de treinamento, agindo (analisando) em várias camadas de dados não lineares para analisar padrões e classificações”. Com a grande ascensão das GPUs, o processamento de várias camadas passou a ser cada vez mais eficiente e rápido, e com essa maior facilidade e, com a queda de custo de hardware, atualmente é umas das áreas mais desenvolvidas dentro das universidades, ainda mais por ser uma técnica aprimorada (que passou a ser explorada) a pouco tempo. Mas apesar de recente, muitas empresas já fazem estudos e desenvolvimentos dentro dessa área. [11]

REDES CONVOLUCIONAIS

A convolução é uma operação matemática que atua sobre duas funções gerando uma terceira, que geralmente é a primeira função modificada ($f(x) * g(x) f'(x)$). [11]

Uma convolução unidimensional discreta, sendo, $x \in \mathbb{Z}$, é definida por [5, 11]:

$$f * g[x] = \sum_{m=-\infty}^{\infty} f[m]g[x - m]$$

“A partir da convolução, em meados de 2006, pesquisadores da CIFAR (Canadian Institute for Advanced Research) adicionaram a esse conceito matemático em uma rede neural artificial, a fim de produzir uma rede que não precisasse de informações pré-rotuladas, assim a máquina possuiria um método de aprendizagem não supervisionada”.

Essa nova técnica combinada às novas GPUs, gerou uma rede neural muito mais confiável e eficiente. [11]

“Para uma rede neural ser convolucional, ela precisa ter uma convolução em pelo menos uma de suas camadas. Em um cenário de reconhecimento de imagens utilizando uma rede neural artificial a convolução funcionaria da seguinte maneira, uma imagem seria dividida em seus pixels, formando assim uma matriz $f(x)$, para analisar essa imagem, é utilizada um filtro $g(x)$, que terá seus pesos invertidos em $g'(x)$, e esse por sua vez, irá passar por toda a matriz efetuando o cálculo convolucional, que ao final de sua execução, irá ter uma nova matriz $f'(x)$, também conhecida como *feature map*”, Figura 8 e 9. [5, 11]

Figura 8. Função, Filtro, e Filtro balanceado para operação. [5]

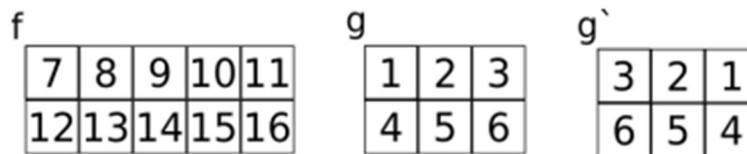
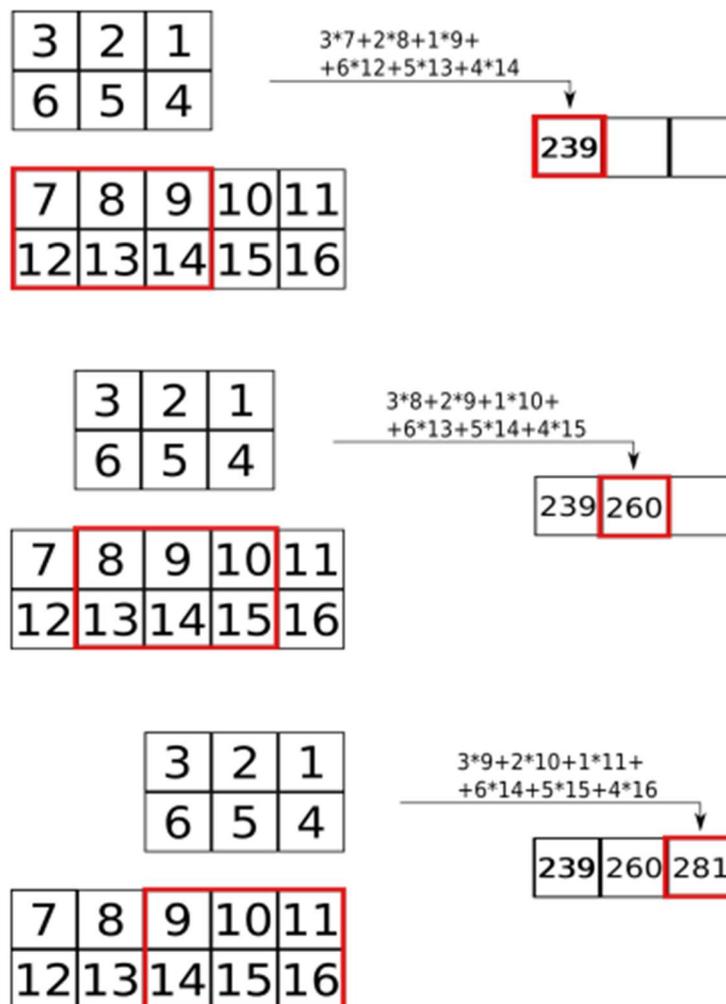


Figura 9. Exemplo funcional de uma convolução em uma matriz. [5]



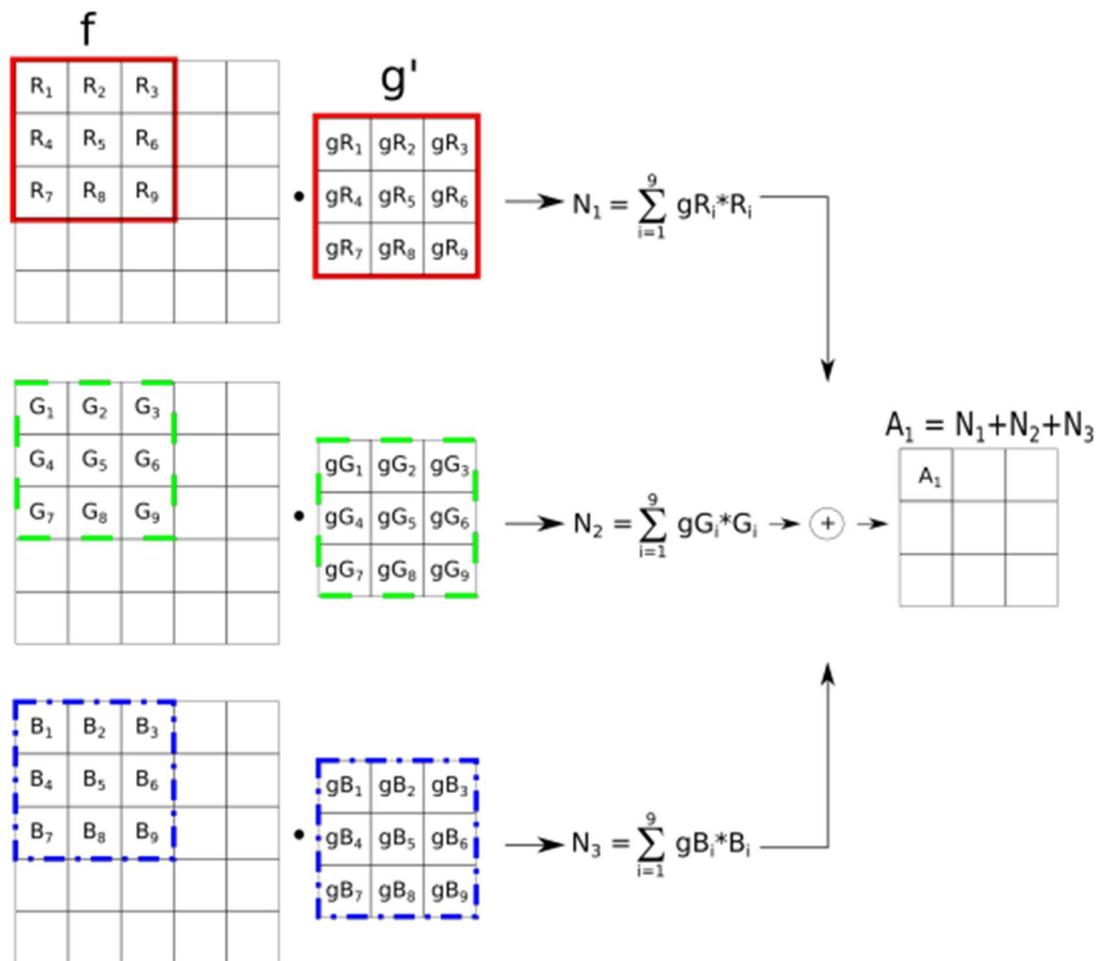
Para o tratamento de imagens ao invés de utilizar a convolução unidimensional, utiliza-se a convolução bidimensional, para fazer análise dos eixos x e y da matriz adquirida na imagem, a equação assume seguinte forma:

$$f * g[x, y] = \sum_{m=-\infty}^{\infty} \sum_{n=-\infty}^{\infty} f[n, m] g[x - n, y - m]$$

[5;11]

Se a imagem analisada estiver em preto e branco, o processo bidimensional acima é o utilizado, mas caso seja uma imagem colorida, o processo passa a ser analisado por 3 filtros, sendo cada um responsável por analisar (bidimensionalmente) uma cor do sistema RGB (Vermelho, Verde, Azul) de cada vez. Ao final da análise, os resultados dos três *feature maps* são somados, gerando apenas um *feature map* RGB para uma próxima análise, conforme exemplo apresentado na Figura 10. [5]

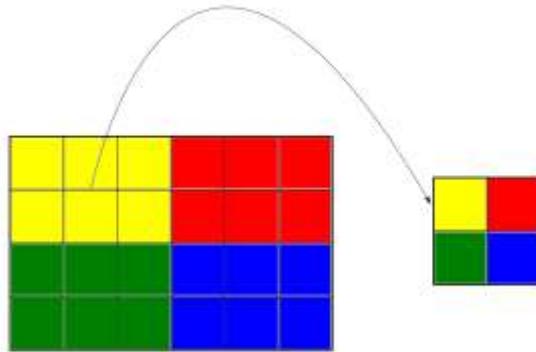
Figura 10. Exemplo funcional de uma convolução em uma matriz de uma imagem colorida, utilizando uma matriz F e G' para cada cor do sistema RGB (Vermelho, Verde, Azul). [5]



POOLING

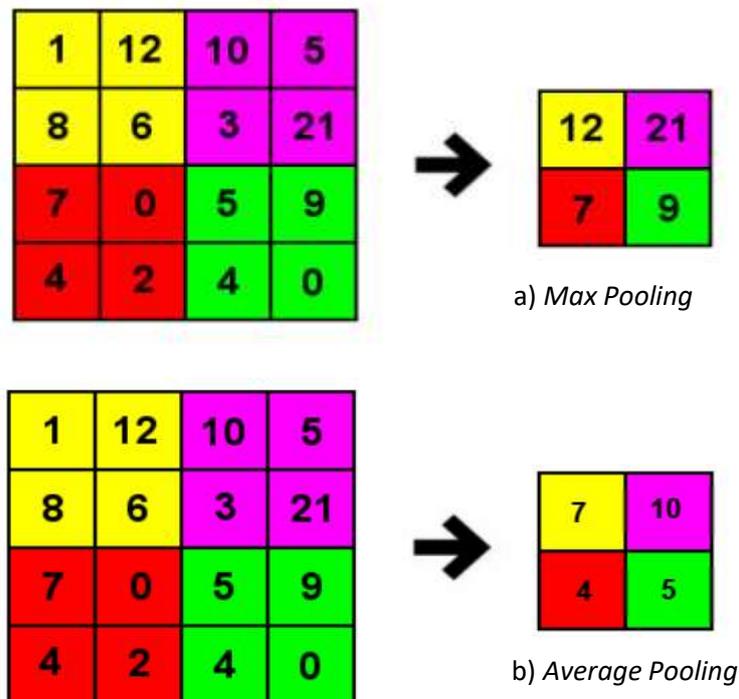
Para haver uma maior agilidade em analisar as imagens, seria preciso que os *feature maps* gerados fossem reduzidos, porém sem perder as informações necessárias, nesse caso, utiliza-se as camadas denominadas *pooling*, Figura 11, reduzindo um *feature map* para um tamanho, que consiga representar a mesma coisa em um ambiente menor. [5;11]

Figura 11. Exemplo básico de como funciona o *Pooling*. [5]



Os dois cálculos mais comuns referentes ao *pooling* são o *Average Pooling*, que calcula a média dos valores da região para a saída; e o *Max Pooling*, que assume o maior valor da região como resultado de saída, como no exemplo da Figura 12. [5]

Figura 12. Exemplo funcional de como funciona o *Max Pooling*, e o *Average Pooling*. [5, 11]



Além de reduzir o tamanho do *feature map* o *pooling* gera a invariância da imagem, que é a ação de representar a parte principal do *feature map*, para a classificação de imagens.

Esse recurso é muito importante, já que ele pode por exemplo, destacar o focinho de um cachorro, já podendo ter um bom argumento para a classificação final. [11]

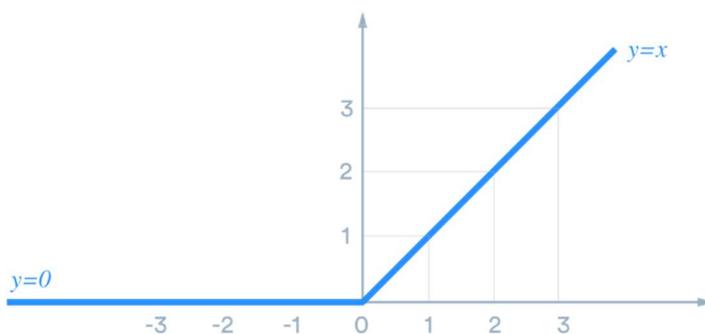
Em geral, as camadas de *pooling* são responsáveis pelo *downsampling* de um *feature map*, geralmente aparecendo após uma camada convolucional, sendo algo indispensável para uma rede neural, pois cria a invariância e diminui o tamanho do *feature map*, reduzindo o número de análises que ela irá executar. Essas duas camadas em sequência, são inspiradas na neurociência visual [5;11]

ATIVAÇÃO

Para a rede obter êxito na sua parte de reconhecimento, é necessário que a mesma se adeque a modelos não lineares, já que ela não irá avaliar sempre a mesma foto, assim ela precisa se adaptar a cada análise, para poder achar o determinado padrão. Como já citado em neurônios artificiais, há uma multiplicação entre camadas seguida de uma função de ativação (podendo ser ou não linear), para cumprir a ação descrita. Se a rede não passar pela função de ativação, ela simplesmente não cumprirá a sua ação. [5]

Apesar de existirem diversas funções de ativação, a principal a ser utilizada na maioria das redes convolucionais é a *ReLU* (*Rectified Linear Unit*) (Figura 13), devido a sua baixa complexidade matemática, e por possuir uma alta eficiência para análises computacionais, fazendo que muitas vezes economize tempo no treinamento da rede. Ela é responsável por ignorar todas as entradas que possuam um resultado inferior a zero, e retorna o valor da função, caso ele seja positivo, ou seja $x < 0 \therefore f(x) = 0, x \geq 0 \therefore f(x) = x$. Atualmente, *ReLU*, é uma das funções mais utilizadas entre as camadas de convolução, principalmente por analisar as evoluções pelas quais máquina passa. [5,12, 18]

Figura 13. Gráfico ilustrativo da função *ReLU* (*Rectified Linear Unit*). [12]

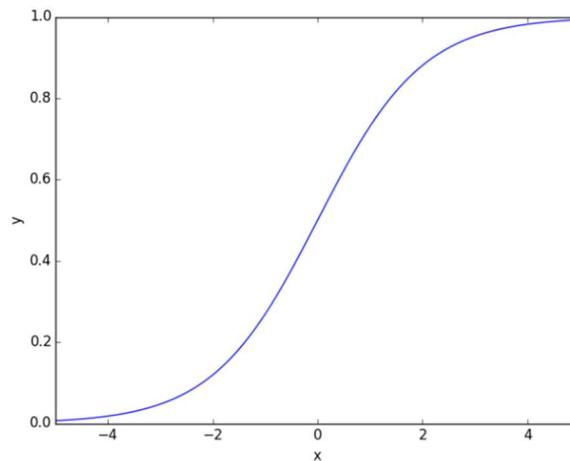


Geralmente ao final das camadas, é aplicada uma ativação diferente, a fim de medir a evolução pela qual a rede passou, medindo desde de a sua perda até aos seus ganhos. Nessa rede, a função final acaba sendo a *sigmoid*, Figura 14, sendo definida como:

$$f(x) = \frac{1}{1+e^{-x}}$$

[5, 14]

Figura 14. Gráfico ilustrativo da função *Sigmoid*. [14]



A sigmoid atuará na fase de reconhecimento, e ela representa uma probabilidade que o valor irá atingir, fazendo assim a previsão da imagem. [14]

REDES PROFUNDAS

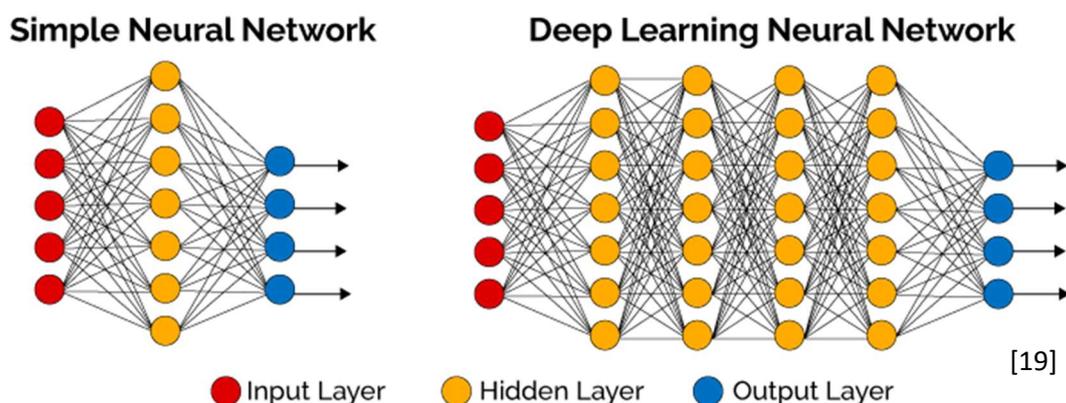
São as redes utilizadas para fazer as várias comparações dos *features maps*, é assim denominada, redes profundas por ter múltiplas camadas ocultas, sendo que nenhuma dessas camadas é uma camada de entrada ou uma camada de saída. Sua funcionalidade é receber os dados de uma determinada camada, e repassar eles para diversas outras, até o final que geralmente termina com uma função que irá classificar o dado. Um exemplo utilizado seria:

$$\begin{aligned}h_1 &= \text{ReLU}(xW_1) \\h_2 &= \text{ReLU}(h_1W_2) \\h_3 &= \text{ReLU}(h_2W_3) \\f(x) &= \text{Sigmoid}(h_3)\end{aligned}$$

[5]

No exemplo, h_1 representa a entrada da camada profunda, nela possui uma função de ativação ReLU, seguida do valor adquirido, e do peso sináptico da camada, no caso W_1 . A rede segue nas camadas h_2 e h_3 , sendo que o valor recebido em cada uma para fazer o cálculo, é o resultado da camada anterior. Até chegar a função final $f(x)$, que irá classificar o resultado da última camada profunda (h_3), nesse caso, foi utilizada a função *sigmoid*. [5]

Figura 15. Gráfico ilustrativo entre uma rede neural artificial e uma rede neural artificial profunda. [19]



DESENVOLVIMENTO

O código a ser analisado foi produzido por *Harrison Kinsley* em “**How to use your trained model: Deep Learning basics with Python, TensorFlow and Keras p.6**” [16]. Baseando nessa publicação, foi escrita uma visão e tradução do material disponibilizado, nele apresenta-se uma rede utilizando *Deep Learning* que consegue realizar a classificação entre cachorros e gatos. Para o desenvolvimento da análise, foi utilizado uma plataforma virtual do ambiente Anaconda, com Python 3.6, TensorFlow 1.13.1, e a biblioteca Keras.

O código fonte está totalmente disponível em [16] na língua inglesa.

BASE DE DADOS

Para poder efetuar a análise foi preciso utilizar uma base de dados (*dataset*) disponibilizada no *download center* da *Microsoft* [17], contendo 25000 imagens de gatos e cachorros com conteúdo e tamanho diferenciável. Dessa forma, foi preciso preparar o *dataset* para o poder utilizar na rede. Nesta aplicação 70% do *dataset* foi utilizado para a base teste, e 30% foi utilizado para a validação.

O primeiro passo é a importação de algumas bibliotecas para efetuar as alterações nas imagens. A biblioteca *numpy* para criação de matrizes, a *matplotlib* gerando gráficos 2D para os futuros resultados do treinamento, o *OS* para trazer as informações de sistemas necessárias para manipulação de arquivos, o *cv2* para manipulação de imagens, e por fim o *tqdm* para obter uma barra de evolução, facilitando a análise visual do processo.

Figura 16. Importações utilizadas para criar o *dataset* (extraído de trechos desenvolvido pelo autor).

```
import numpy as np
import matplotlib.pyplot as plt
import os
import cv2
from tqdm import tqdm
```

Figura 17. Imagem original utilizada (extraído do *dataset* disponível em [17])



Figura 18. Conversão de imagem em matriz, e de colorida para preto e branco (extraído de trechos desenvolvido pelo autor).

```
DATADIR = r"E:\tcc\PetImages"
CATEGORIES = ["Dog", "Cat"]

for category in CATEGORIES: # criando categorias
    path = os.path.join(DATADIR,category) # criando caminho para imagens
    for img in os.listdir(path): # passagem das imagens
        img_array = cv2.imread(os.path.join(path,img) ,cv2.IMREAD_GRAYSCALE) # convertendo para uma matriz
        plt.imshow(img_array, cmap='gray') # criando a imagem na matriz
        plt.show() # mostrando

        break # mostrando somente uma imagem
    break
```

```
print(img_array)
[[117 117 119 ... 133 132 132]
 [118 117 119 ... 135 134 134]
 [119 118 120 ... 137 136 136]
 ...
 [ 79  74  73 ...  80  76  73]
 [ 78  72  69 ...  72  73  74]
 [ 74  71  70 ...  75  73  71]]

print(img_array.shape)
(375, 500)
```

Em seguida, a mesma imagem tem seu tamanho alterado para o padrão que será utilizado no projeto. Para essa análise, todas as imagens possuem o tamanho 50x50 pixels.

Figura 21. Embaralhando a ordem das imagens (extraído de trechos desenvolvido pelo autor).

```
import random

random.shuffle(training_data)

X = []
y = []

for features,label in training_data:
    X.append(features)
    y.append(label)

print(X[0].reshape(-1, IMG_SIZE, IMG_SIZE, 1))

X = np.array(X).reshape(-1, IMG_SIZE, IMG_SIZE, 1)

[[[[200]
   [ 62]
   [176]
   ...
   [ 56]
   [ 53]
   [ 57]]

  [[205]
   [198]
   [ 68]
   ...
   [ 62]
   [ 63]
   [ 64]]

  [[204]
   [203]
   [ 77]
   ...
   [ 70]
   [ 93]
   [ 96]]

  ...
```

Figura 22. Salvar o *dataset* (extraído de trechos desenvolvido pelo autor).

```
import pickle

pickle_out = open("X.pickle","wb")
pickle.dump(X, pickle_out)
pickle_out.close()

pickle_out = open("y.pickle","wb")
pickle.dump(y, pickle_out)
pickle_out.close()
```

DESENVOLVIMENTO DE UMA APLICAÇÃO DE REDE NEURAL

Nesta seção mostramos como desenvolver uma aplicação de Aprendizado de Máquina utilizando *TensorFlow*, a biblioteca *Keras* em *Python*. Como na criação do *dataset*, inicialmente se faz os *imports* das bibliotecas utilizadas na rede neural, a maioria das bibliotecas a serem utilizadas foram retiradas do *Keras*, que é uma API de redes neurais, utilizada no *TensorFlow* (biblioteca focada em aprendizado de máquinas desenvolvida pelo *Google*). Os modelos utilizados nesse projeto foram: *Sequential*, para utilizar de seu modelo de rede pré-construído, e adicionar as camadas de densidade (*Dense*) para prever os rótulos, a camada responsável pela ativação do neurônio (*Activation*), a camada *Flatten*, utilizada para transformar a matriz em vetor, a camada de convolução (*Conv2D*) que efetuará a convolução nos dados processados, e pôr fim a camada de *pooling* (*MaxPooling2D*) para reduzir os dados analisados. Além dos *imports* relacionados a rede neural, foi adicionado *pickle* novamente para salvar ao final da execução, *time* para verificação de tempo, e o *TensorBoard* para ver os resultados da rede em forma gráfica. Apresentando a chamada de *imports* na Figura 23.

Figura 23. *Imports* para construção da rede neural profunda (extraído de trechos desenvolvido pelo autor).

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Activation, Flatten
from tensorflow.keras.layers import Conv2D, MaxPooling2D
from tensorflow.keras.callbacks import TensorBoard
import pickle
import time
```

Após os *imports*, são carregados os *datasets* criados com o *pickle* e em seguida o *dataset* é passado por uma normalização de seus dados. Para isso, realiza a operação $\frac{x}{255.0}$, pois o valor máximo que o pixel poderá possuir é 255, logo, é feita essa divisão para se obter o valor normalizado do dado. É declarado depois o número de camadas densas, o tamanho de cada camada (densa e convolucional), e o número de camadas convolucionais. Visualização de como é realizada a importação na Figura 24.

Figura 24. Importação e normalização dos *datasets*, e atribuição de número e tamanho das camadas densas e convolucionais (extraído de trechos desenvolvido pelo autor).

```
pickle_in = open("X.pickle","rb") #importação dos datasets
X = pickle.load(pickle_in)

pickle_in = open("y.pickle","rb")
y = pickle.load(pickle_in)

X = X/255.0 #normalização do dataset

dense_layers = [0, 1, 2] #número de camadas densas
layer_sizes = [32, 64, 128] #tamanho da camda densa
conv_layers = [1, 2, 3] #número de camadas convolucionais
```

Inicia-se a criação da rede neural, cria-se uma estrutura de repetição baseada nos valores atribuídos na Figura 24, e para não confundir os resultados gerados, cada análise é gravada com o nome dos dados da estrutura de repetição. Para iniciar a rede neural, é utilizado o modelo *Sequential* para criar a estrutura, a camada de entrada será uma camada convolucional, com o número de camadas de *layer_size*, um filtro 3x3, e uma matriz de entrada do *dataset*. Terminada, os dados são passados pela camada de ativação (sendo utilizada a função *ReLU*), e por fim a camada *pooling*, utilizando o método de *max-pooling* (*MaxPooling2D*) para obter os resultados relevantes da matriz analisada, apresentada na Figura 25.

Na sequência, todo o processo é refeito (convolução, ativação e *pooling*), porém com a diminuição de uma camada convolucional conforme Figura 25.

Passando por essas duas etapas, as informações adquiridas na matriz são transformadas em um vetor, e repassadas para análise nas camadas densas, após o processamento, é passada pela ativação, novamente utilizando a função *ReLU*. A análise final é passada por uma camada densa, e a ativação que irá classificar a imagem, para essa ação a função a ser utilizada será a *Sigmoid*. Para analisar o resultado final, é chamado o *TensorBoard* para análise gráfica, apresentada na Figura 26.

Figura 25. Rede neural convolucional densa (extraído de trechos desenvolvido pelo autor).

```
for dense_layer in dense_layers:
    for layer_size in layer_sizes:
        for conv_layer in conv_layers:
            NAME = "{}-conv-{}-nodes-{}-dense-{}".format(conv_layer, layer_size, dense_layer, int(time.time()))
            print(NAME)

            model = Sequential() #estrutura da rede neural

            model.add(Conv2D(layer_size, (3, 3), input_shape=X.shape[1:])) #rede convolucional recebendo a primeira imagem
            model.add(Activation('relu')) #camada de ativação
            model.add(MaxPooling2D(pool_size=(2, 2))) #camada pooling

            for l in range(conv_layer-1):
                model.add(Conv2D(layer_size, (3, 3))) #rede convolucional recebendo dados da outra rede
                model.add(Activation('relu')) #camada de ativação
                model.add(MaxPooling2D(pool_size=(2, 2))) #camada pooling

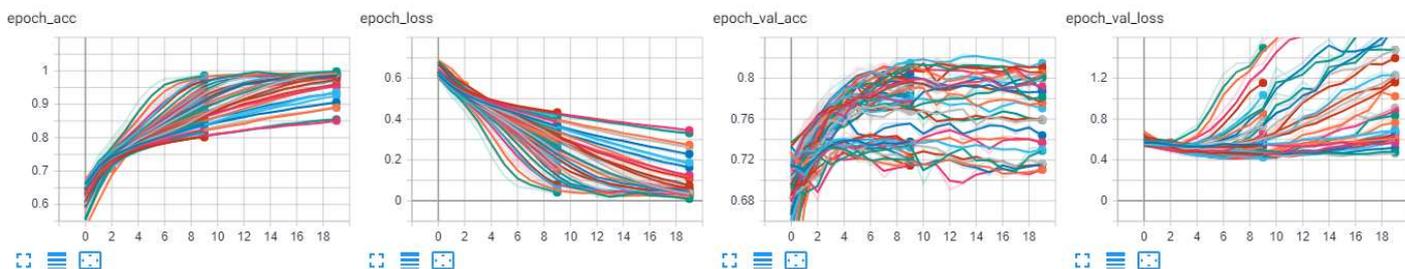
            model.add(Flatten()) #transformação de matriz em vetor

            for _ in range(dense_layer):
                model.add(Dense(layer_size)) #camada densa com o numero atribuido em layer_size
                model.add(Activation('relu')) #camada de ativação

            model.add(Dense(1)) #camada densa com uma camada
            model.add(Activation('sigmoid')) #camada de ativação para classificar

            tensorboard = TensorBoard(log_dir="logs/{}".format(NAME))
```

Figura 26. Resultado gráfico através do TensorBoard (extraído de trechos desenvolvido pelo autor).



Terminada a rede, ela é compilada através de uma função de perda (*binary_crossentropy*), um otimizador (*adam*), e a métrica de acerto (*accuracy*). Após a compilação, é feito um ajuste através do método *fit*, e treina o modelo em um determinado número de épocas (*epochs*). E então os resultados começam a aparecer na tela. Na Figura 27 representa esta parte do código, e a Figura 28 exibe uma pequena parte dos resultados gerados.

Figura 27. Compilação e ajustamento da rede, para então executar e mostrar os resultados (extraído de trechos desenvolvido pelo autor).

```
model.compile(loss='binary_crossentropy',
              optimizer='adam',
              metrics=['accuracy'],
              )

model.fit(X, y,
        batch_size=32,
        epochs=20,
        validation_split=0.3,
        callbacks=[tensorboard])

model.save('64x3-CNN.model')
```

Figura 28. Parte dos resultados gerados (1 convolução, 32 nós, 0 densas, 20 épocas) (extraído de trechos desenvolvido pelo autor).

```
1-conv-32-nodes-0-dense-1574614753
Train on 17462 samples, validate on 7484 samples
Epoch 1/20
17462/17462 [=====] - 8s 457us/sample - loss: 0.6183 - acc: 0.6621 - val_loss: 0.5711 - val_acc: 0.7104
Epoch 2/20
17462/17462 [=====] - 8s 446us/sample - loss: 0.5513 - acc: 0.7239 - val_loss: 0.5699 - val_acc: 0.7022
Epoch 3/20
17462/17462 [=====] - 8s 469us/sample - loss: 0.5200 - acc: 0.7444 - val_loss: 0.5919 - val_acc: 0.6808
Epoch 4/20
17462/17462 [=====] - 8s 448us/sample - loss: 0.4983 - acc: 0.7608 - val_loss: 0.5486 - val_acc: 0.71247
Epoch 5/20
17462/17462 [=====] - 8s 446us/sample - loss: 0.4800 - acc: 0.7774 - val_loss: 0.5318 - val_acc: 0.7401
Epoch 6/20
17462/17462 [=====] - 8s 449us/sample - loss: 0.4622 - acc: 0.7844 - val_loss: 0.5276 - val_acc: 0.7466
```

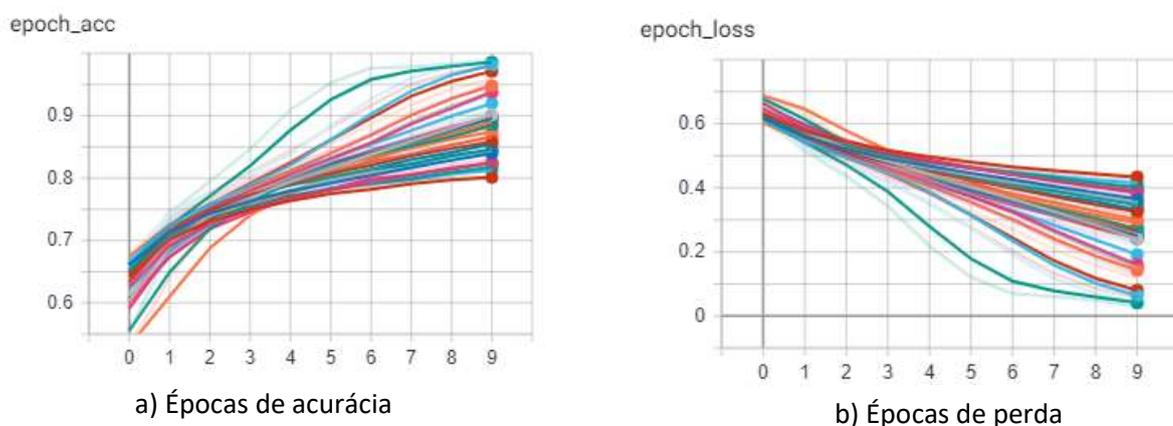
TESTES E RESULTADOS

O teste busca analisar qual o melhor resultado entre várias redes neurais com diversas características diferentes como: número de camadas densas, número de camadas convolucionais, número de nós por camadas e, também, qual o número de épocas para obtenção dos resultados. Como o material utilizado para esse trabalho tem uma função didática, o autor analisa várias redes neurais, com a intenção de mostrar como cada camada pode ter uma determinada influência na rede neural. Para mostrar essas diferenças ele trabalhou com modelos possuindo de zero a duas camadas densas, e de uma a três camadas convolucionais, ambas possuem nós de 32, 64, ou 128 camadas internas, possuindo 10 épocas (0 a 9) por análise. Lembrando que o *dataset* utilizado está dividido em 70% das imagens para aprendizado (17500 imagens) e 30% das imagens para validação (7500 imagens). Dadas as características, as redes analisadas estão descritas na Tabela 1. Os resultados dos testes estão expressos na Figura 29.

Tabela 1. Estruturas das redes criadas por [16] para análise.

Convolução	Nós	Densa
1 conv	32 nodes	0 dense
1 conv	64 nodes	0 dense
1 conv	128 nodes	0 dense
1 conv	32 nodes	1 dense
1 conv	64 nodes	1 dense
1 conv	128 nodes	1 dense
1 conv	32 nodes	2 dense
1 conv	64 nodes	2 dense
1 conv	128 nodes	2 dense
2 conv	32 nodes	0 dense
2 conv	64 nodes	0 dense
2 conv	128 nodes	0 dense
2 conv	32 nodes	1 dense
2 conv	64 nodes	1 dense
2 conv	128 nodes	1 dense
2 conv	32 nodes	2 dense
2 conv	64 nodes	2 dense
2 conv	128 nodes	2 dense
3 conv	32 nodes	0 dense
3 conv	64 nodes	0 dense
3 conv	128 nodes	0 dense
3 conv	32 nodes	1 dense
3 conv	64 nodes	1 dense
3 conv	128 nodes	1 dense
3 conv	32 nodes	2 dense
3 conv	64 nodes	2 dense
3 conv	128 nodes	2 dense

Figura 29. Resultado dos testes efetuados por [16] (extraído de trechos desenvolvido pelo autor).



Analisando os gráficos de teste, é possível perceber que a rede é funcional devido aos seus valores de acurácia (Figura 29 a) estarem entre 0.8 e um pouco acima de 1.0; e os valores de perda (Figura 29 b) estarem abaixo de 0.5. Esses são bons valores, pois a

acurácia mostra que em teste a máquina chegou a pelo menos 80% de aprendizado sobre as imagens e, o valor de perda, significa que sua perda de informação adquirida é de menos de 50%. O resultado poderia ser negativo caso a acurácia fosse baixa (menor que 70%), ou caso a taxa de perda fosse muito alta (acima de 50%), mesmo tendo uma taxa de acurácia alta, pois não adianta a máquina aprender, e perder seus dados em sequência. Apesar dos bons resultados, existem alguns modelos que se destacam mais que os outros nessa fase de testes. Os três principais (Figura 30) a serem utilizados foram os que possuíam uma convolução, 64 nós (vermelha) e 128 nós (azul e verde), e uma camada densa (azul), ou duas (vermelha, verde). Dando um principal destaque para a rede com uma camada convolucional, 128 nós, e duas camadas densas (verde).

Figura 30. Melhores resultados dos testes efetuados por [16] (extraído de trechos desenvolvido pelo autor).

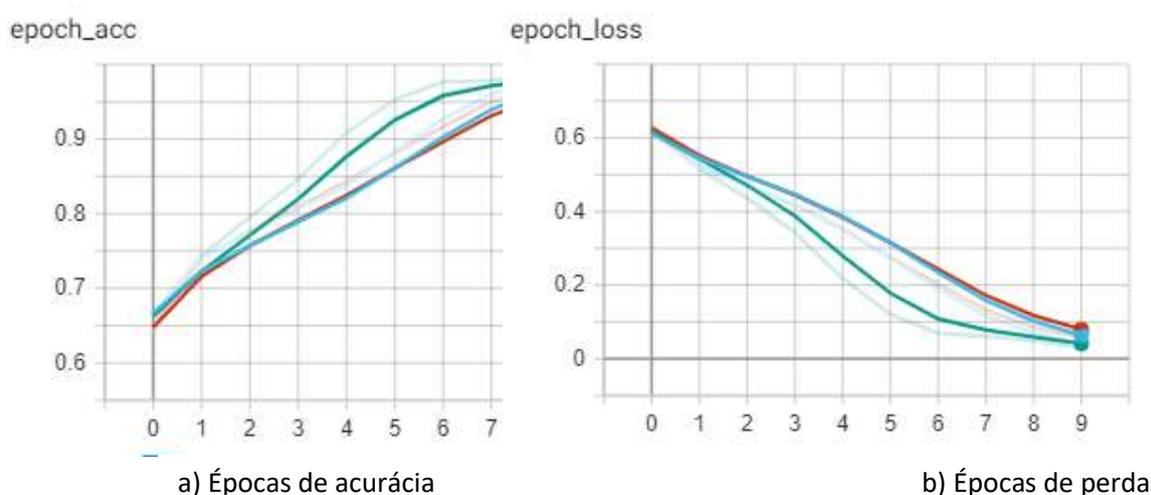


Tabela 2. Resultados dos testes do autor

Modelo	Aprendizado	Perda	Tempo'	Cor
1-conv-128-nodes-1-dense	0.9767	0.0774	~13.54	Azul
1-conv-64-nodes-2-dense	0.9765	0.0670	~4.52	Vermelho
1-conv-128-nodes-2-dense	0.9871	0.0398	~13.44	Verde

Porém ao analisar as suas validações, percebe-se que a inteligência não atingiu a mesma eficiência que nos testes. A rede 1-conv-128-nodes-1-dense, até desempenha um bom papel no aprendizado, porém tem uma taxa de perda muito alta, já que a mesma tem que chegar o mais próximo de 0, como os exemplos contidos na Tabela 2. A Figura 31, e a Tabela 3 mostram o resultado negativo que esses três modelos obtiveram ao validar as imagens.

Figura 31. Resultados das validações efetuadas por [16], referente a tabela 1 (extraído de trechos desenvolvido pelo autor).

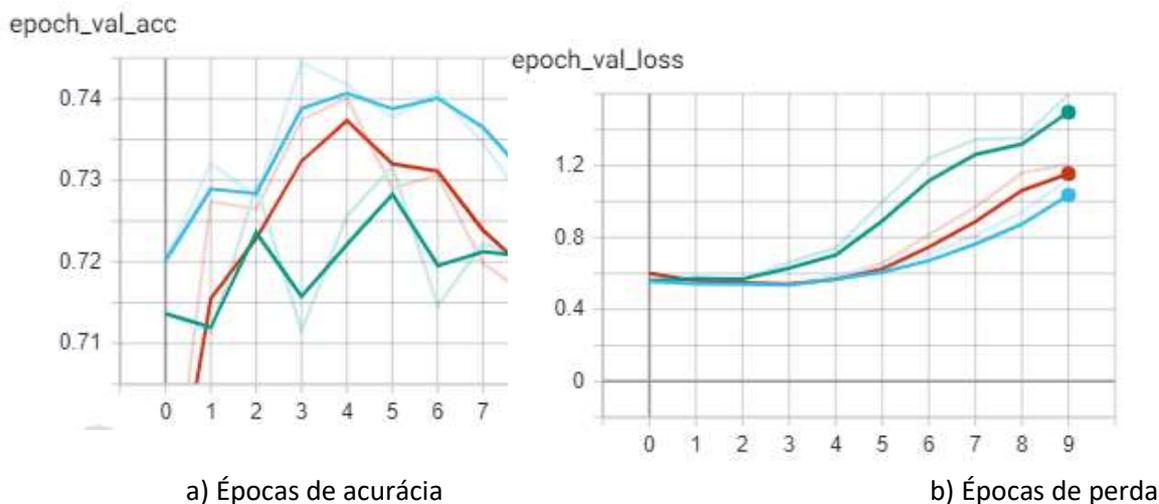


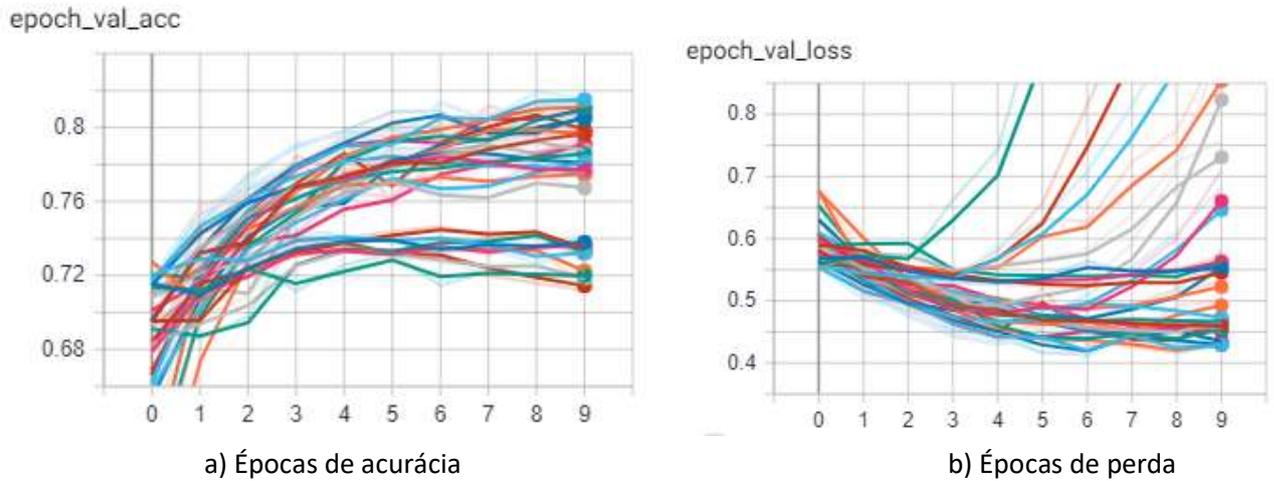
Tabela 3. Resultados das validações do autor

Modelo	Aprendizado	Perda	Tempo'	Cor
1-conv-128-nodes-1-dense	0.7356	0.9490	~13.54	Azul
1-conv-64-nodes-2-dense	0.7159	1.3458	~4.52	Vermelho
1-conv-128-nodes-2-dense	0.7218	1.5668	~13.44	Verde

É normal os resultados da validação serem um pouco piores do que os dos testes, pois não está em um ambiente de teste, sendo assim ela trata os dados de uma forma mais aprofundada do que em uma zona de testes, em que pequenas diferenças são ignoradas para um funcionamento mais rápido. Porém os casos na Tabela 3, apesar de terem valores aceitáveis de acurácia (acima de 70% nesse caso), a variação da taxa de perda aumentou muito chegando a ultrapassar o valor 1.0 (nesse caso mais de 100% das informações adquiridas no processamento de imagens foi misturada), nessas validações, as redes não conseguiriam fazer as distinções, já que as mesmas misturaram as informações adquiridas, analisando de uma forma geral, nesses casos apresentados, as chances de ela acertar se a imagem é de gato ou cachorro, seria na base de chute. Sendo assim, nem sempre o que é adquirido nos treinos é realizado na validação.

Vamos analisar as validações das redes neurais citadas no começo do capítulo, que foram propostas por [16] na Figura 32.

Figura 32. Resultado das validações efetuadas por [16] (extraído de trechos desenvolvido pelo autor).



Os modelos funcionais apresentados nos gráficos (Figura 32) são os que se possuem um valor de acurácia (Figura 32 a) acima de 0.7 (nesse caso todas se encaixam), e estejam com um valor de perda (Figura 32 b) abaixo de 0.5, para ter certeza de que o modelo apresentado possua a capacidade de aprender, e manter as informações adquiridas. Analisando visualmente o gráfico, e eliminando os modelos que visualmente não atendem os pré requisitos, chegamos a três soluções (Figura 33), que mostram as melhores características para poder executar essa operação.

Figura 33. Resultado das validações com menor taxa de perda efetuadas por [16] (extraído de trechos desenvolvido pelo autor).

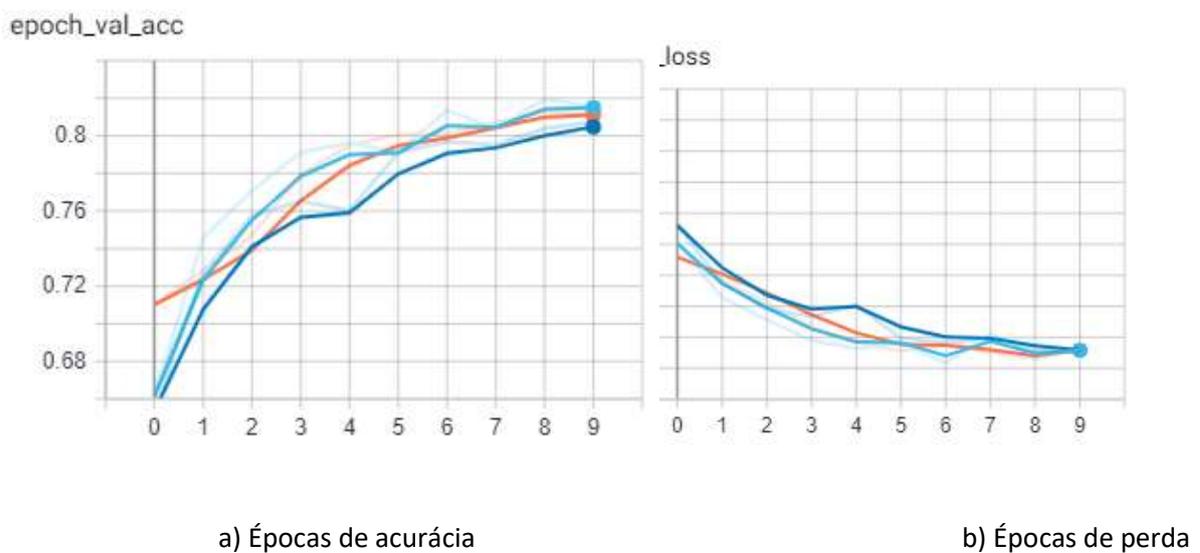


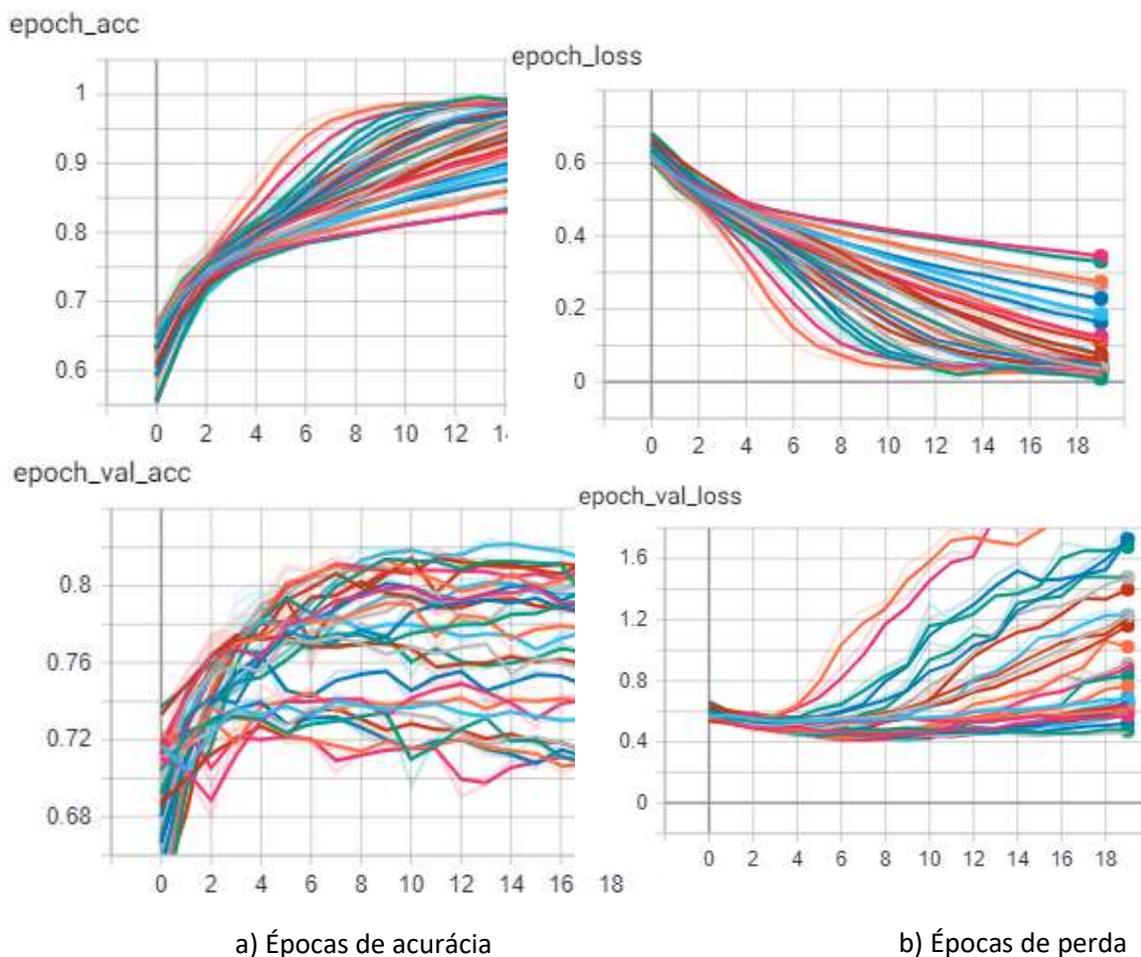
Tabela 4. Resultado das validações com menor taxa de perda efetuadas pelo autor

Modelo	Aprendizado	Perda	Tempo'	Cor
3-conv-128-nodes-0-dense	0.8180	0.4360	~14.65	Azul Claro
3-conv-64-nodes-0-dense	0.8112	0.4467	~6.4	Laranja
3-conv-32-nodes-2-dense	0.7981	0.4450	~3.17	Azul Escuro

Analisando a Tabela 4, percebe-se que o número de convoluções faz uma grande diferença na validação da rede, já que os modelos que apresentaram as melhores validações possuem três camadas convolucionais, o que gerou a melhoria na taxa de perda, e até mesmo na taxa de aprendizado da máquina. Isso ocorre pelo maior tratamento dos dados ao serem analisados nas camadas convolucionais adicionadas.

Baseando-se nessas redes, foram adicionadas mais dez épocas (0 a 19), para ver a variação que a rede pode obter, e se as taxas de aprendizado e perda iriam obter uma melhoria, já que os dados seriam analisados por um período maior.

Figura 34. Resultado dos testes e validações efetuadas com 20 épocas (extraído de trechos desenvolvido pelo autor).



Visualmente, o aumento de épocas (Figura 34) foi negativo para os modelos de redes neurais. Nas validações, o valor de acurácia ficou acima de 0.7 (Figura 34 a), porém os valores de perda (Figura 34 b) se mostram maiores do que a versão com 10 épocas (Figura 32 b). No gráfico de perda (Figura 34 b) é evidente de que as redes neurais ou mantem um padrão desde o início, ou senão, começam a ter variações após um tempo de análise, que fazem com que o valor da perda aumente, o que é indesejável para uma rede neural. Muito provavelmente, isso ocorre devido a padronização encontrada em sequência, o que pode acabar confundindo a máquina.

Para uma melhor comparação, as melhores estruturas encontradas com dez épocas (Figura 31 e Tabela 4), serão comparadas com as de 20 épocas (Figura 35 e Tabela 5), com relação a ativação delas, já que nos gráficos de teste (Figura 34), os resultados apresentados são superiores aos dados aprendizados nos gráficos de testes com 10 épocas (Figura 29).

Figura 35. Resultado das validações efetuadas com 20 épocas com as estruturas da tabela 4 (extraído de trechos desenvolvido pelo autor).

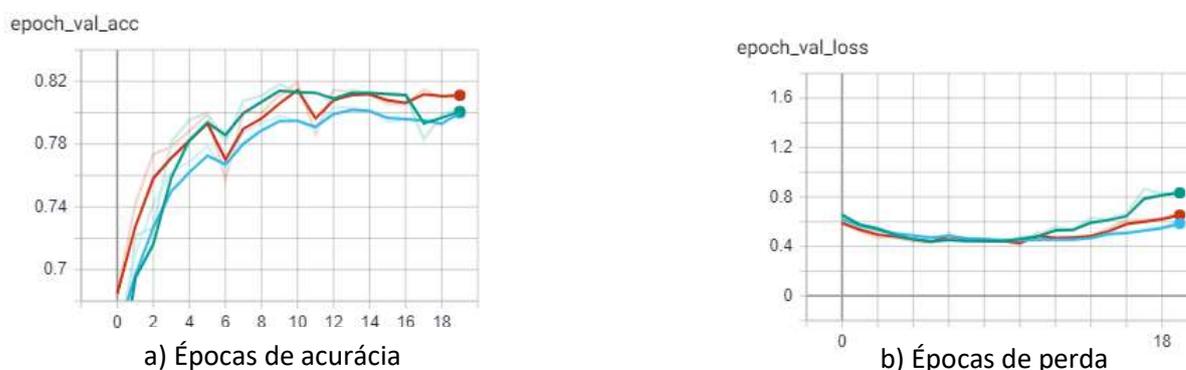


Tabela 5. Resultado das validações com base na tabela 3, possuindo 20 épocas.

Modelo	Aprendizado	Perda	Tempo'	Cor
3-conv-128-nodes-0-dense	0.8030	0.8443	~29.29	Verde
3-conv-64-nodes-0-dense	0.8113	0.6729	~13.00	Vermelho
3-conv-32-nodes-2-dense	0.8038	0.6046	~6.34	Azul

Comparando as tabelas 4 e 5, é nítido que os três casos regrediram em melhorar suas taxas, ao ponto de que o modelo 3-conv-128-nodes-0-dense com 10 épocas possuía o melhor resultado, enquanto o mesmo modelo com 20 épocas tem sua taxa de perda aumentada de 0.4360 para 0.8443.

Executando a mesma análise de gráficos da Figura 32, e eliminado as redes com os maiores valores de perda, chegou-se a um novo modelo de rede neural, sendo esse o que possuía a menor taxa de perda, chegando mais perto do desejável, sendo ele composto por 3-conv-32-nodes-0-dense. Sua representação está na Figura 36.

Figura 36. Resultado dos testes e validações efetuadas com o modelo 3-conv-32-nodes-0-dense com 20 épocas (extraído de trechos desenvolvido pelo autor).

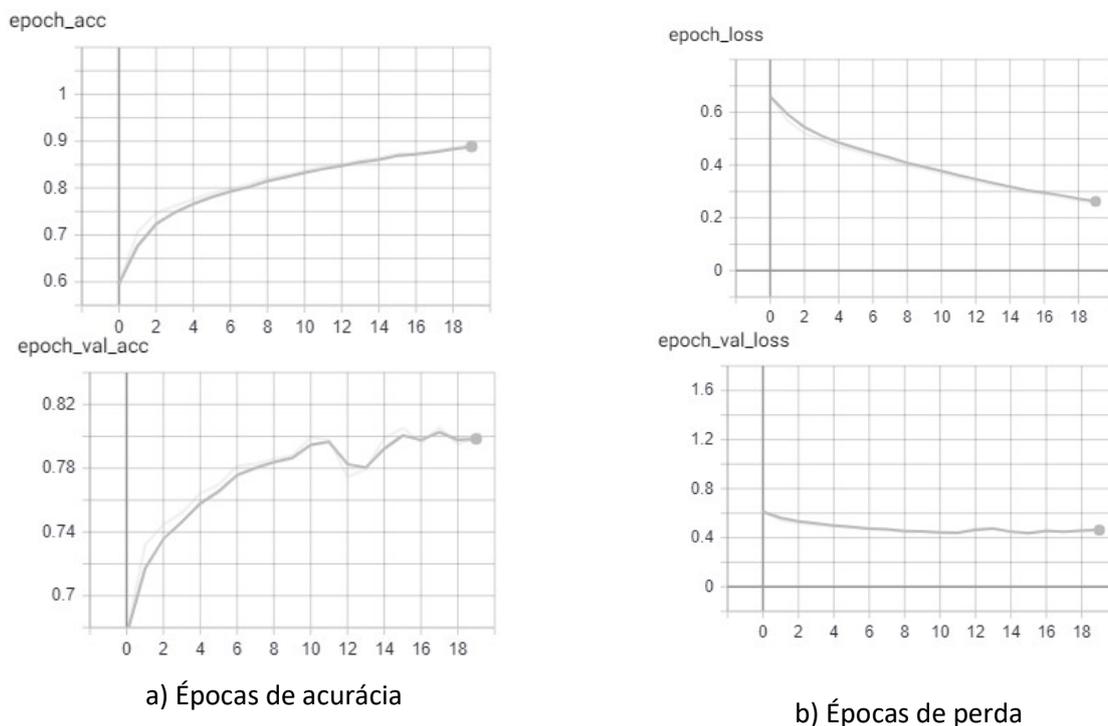


Tabela 6. Resultado das validações efetuadas com o modelo 3-conv-32-nodes-0-dense com 20 épocas.

Modelo	Aprendizado	Perda	Tempo'	Cor
3-conv-32-nodes-0-dense	0.7988	0.4640	~6.35	Cinza

Se compararmos esse modelo (Figura 36 e Tabela 6), com os modelos da Tabela 5, percebe-se a grande diferença em relação a taxa de perda, que como já citado, é a principal forma de saber se a inteligência obtem progresso ou não, por não perder as informações adquiridas.

Parando para analisar os números das Tabelas 4, 5 e 6, é notável que em todas as estruturas de melhor taxa possuem 3 camadas de convolução. Na sua grande maioria, não utiliza as camadas densas (salvo 3-conv-32-nodes-2-dense, que utiliza somente 32 nós, o que evita o conflito de informação que deve estar tendo dentro das camadas). Referente aos nós, em dez épocas, sua variação não é tanto perceptiva, ainda mais que a rede que obteve a melhor taxa de perda (0.4360), possuía 128 nós; porém ao aumentar o número de épocas e estendendo o processo, percebe-se que quanto maior o número de nós, maior a chance da rede neural se perder na hora da validação de dados (totalmente notável na mudança de 0.4360 para 0.8443 no modelo 3-conv-128-nodes-0-dense). Tudo fica mais aparente, quando o melhor modelo com 20 épocas é dado como 3-conv-32-nodes-0-dense, mostrando que trabalhando com um menor número de nós, ele consegue manter uma constante na sua taxa de perda.

Baseando-se nesses fatos e nos trabalhos de [5] e [7], um novo teste foi sugerido. Baseando nos dados já presentes, os modelos a serem testados irão possuir três camadas convolucionais, tendo entre 32, 64, e 128 nós, afim de achar alguma diferença entre eles, e possuíram zero ou três camadas densas, afim de analisar como a camada densa interfere no processo. Como já testado, o número de épocas influencia muito no resultado final, e isso provavelmente se deve a repetição de processos, que ao fazer várias vezes o mesmo caminho, acaba se perdendo na hora da validação. Para resolver esse problema, será adicionado duas camadas de *dropout*, conforme descrito em [5] e [7]. A camada de *dropout* será responsável por eliminar nós aleatoriamente, a cada vez que a informação passar por uma camada; assim pretende se criar uma variação entre as épocas, fazendo com que com a sua taxa de perda na validação se mantenha baixa. E para reanalisar a questão de extensão de épocas, foram adicionas mais 20, sendo assim, esse novo teste será realizado por 40 épocas (0 a 39).

Para as realizações de cima, as alterações feitas no código, serão representadas na Figura 37.

A primeira camada de dropout, é de 0.25, significando que 25% dos nós são eliminados por passagem, e a seguinte é de 0.5, representado 50% dos nós.

Figura 37. Alterações para inclusão da camada *dropout* (extraído de trechos desenvolvido pelo autor).

```
from tensorflow.keras.models import Sequential
from tensorflow.keras.layers import Dense, Dropout, Activation, Flatten
from tensorflow.keras.layers import Conv2D, MaxPooling2D
from tensorflow.keras.callbacks import TensorBoard
import pickle
import time

pickle_in = open("x.pickle","rb")
X = pickle.load(pickle_in)

pickle_in = open("y.pickle","rb")
y = pickle.load(pickle_in)

X = X/255.0

dense_layers = [0,3]
layer_sizes = [32, 64, 128]
conv_layers = [3]

for dense_layer in dense_layers:
    for layer_size in layer_sizes:
        for conv_layer in conv_layers:
            NAME = "{}-conv-{}-nodes-{}-dense-{}".format(conv_layer, layer_size, dense_layer, int(time.time()))
            print(NAME)

            model = Sequential()

            model.add(Conv2D(layer_size, (3, 3), input_shape=X.shape[1:]))
            model.add(Activation('relu'))
            model.add(MaxPooling2D(pool_size=(2, 2)))

            for l in range(conv_layer-1):
                model.add(Conv2D(layer_size, (3, 3)))
                model.add(Activation('relu'))
                model.add(MaxPooling2D(pool_size=(2, 2)))
                model.add(Dropout(0.25))

            model.add(Flatten())

            for _ in range(dense_layer):
                model.add(Dense(layer_size))
                model.add(Activation('relu'))
                model.add(Dropout(0.5))

            model.add(Dense(1))
            model.add(Activation('sigmoid'))

            tensorboard = TensorBoard(log_dir="logs/{}".format(NAME))

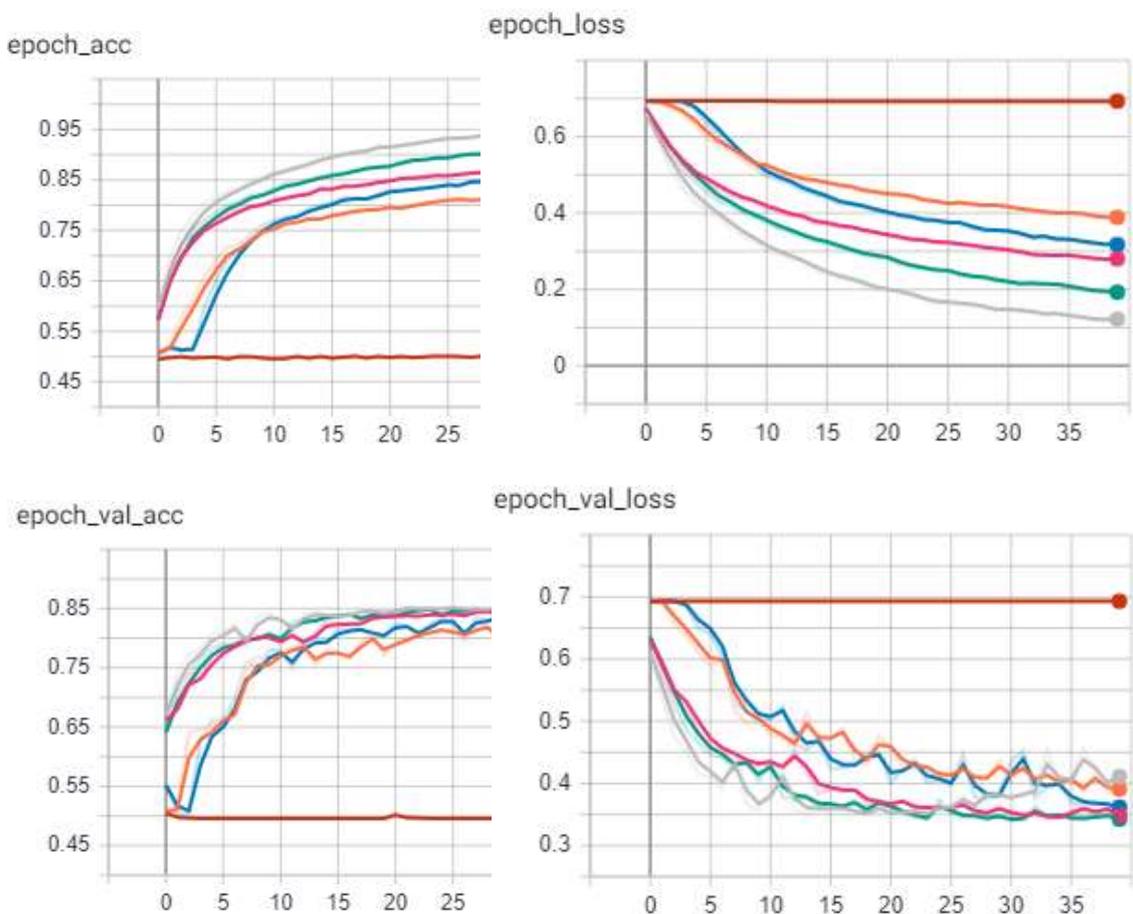
            model.compile(loss='binary_crossentropy',
                          optimizer='adam',
                          metrics=['accuracy'],
                          )

            model.fit(X, y,
                    batch_size=32,
                    epochs=40,
                    validation_split=0.3,
                    callbacks=[tensorboard])

model.save('64x3-CNN.model')
```

Após as alterações os resultados se mostraram bastante positivos, confirmando a importância de se gerar variações em uma rede muito loga. Os resultados estão expostos na Figura 38.

Figura 38. Resultado dos testes e validações efetuadas com 40 épocas, e duas camadas *dropout* (extraído de trechos desenvolvido pelo autor).



a) Épocas de acurácia

b) Épocas de perda

Analisando o gráfico, percebe-se a melhoria obtida na maior parte dos modelos, tendo somente um que não atingiu as expectativas (3-conv-128-nodes-3-dense), sendo assim, utilização do dropout se mostrar essencial para casos que possua várias épocas.

Tabela 7. Resultado das validações efetuadas com 40 épocas, e duas camadas *dropout*.

Modelo	Aprendizado	Perda	Tempo'	Cor
3-conv-128-nodes-0-dense	0.8326	0.4167	~60.00	Cinza
3-conv-64-nodes-0-dense	0.8545	0.3392	~26.70	Verde
3-conv-32-nodes-0-dense	0.8506	0.3430	~12.70	Rosa
3-conv-64-nodes-3-dense	0.8435	0.3590	~30.00	Azul
3-conv-32-nodes-3-dense	0.8252	0.3889	~13.4	Laranja
3-conv-128-nodes-3-dense	0.6932	0.4956	~58.00	Vermelho

Ao analisar os dados da Tabela 7, é nótavel a melhoria obtida principalmente em 3-conv-64-nodes-0-dense, tendo sua taxa de perda em 0.3392, marca que nenhuma das outras redes (tabelas 4, 5 e 6) conseguiram atingir antes. Mostrando que uma boa base para diferenciar duas espécies, seria uma de 3 camadas convolucionais, possuindo 32 ou 64 nós,

utilizando duas camadas de dropout (0.25 e 0.5) e nenhuma camada densa, para evitar conflito de informações.

CONCLUSÕES

No trabalho foi desenvolvido uma rede neural artificial, que tem a capacidade de distinguir cachorros de gatos, sua base foi retirada de [16], ele já estava funcional e cumpria o que havia sido proposto, porém em uma tentativa de analisar seu comportamento com épocas maiores, percebeu-se uma inconstância no código, que fazia com que ele se perdesse; e após algumas análises em [5] e [7], algumas alterações foram feitas no código original. Essas alterações proporcionaram uma melhoria, consertando o problema encontrado no código.

Inicialmente o código continha estruturas que possuíam de zero a duas camadas profundas, de uma a três camadas convolucionais, e nós nos tamanhos de 32, 64 e 128, duas camadas de *pooling*, e três ativações, sendo duas *ReLU* e uma *Sigmoid*, atuando em 10 épocas. Para fazer a análise da rede foi utilizado um *dataset* de 25000 imagens. Na primeira análise, percebeu-se a grande importância de se possuir camadas convolucionais, e que se ao utilizar as redes densas para esses tipos de casos de diferenciação de imagens na rede neural podem acabar se atrapalhando e misturando as informações necessárias para fazer a validação das imagens. Em uma segunda análise, a rede foi colocada a interagir em 20 épocas, para ver seu desempenho, que acabou obtendo mal resultados. Por fim com a base utilizada em [5] e [7], duas camadas de *dropout* foram adicionadas a rede, o que fez os dados que antes se perdiam, tomassem caminhos diferentes, assim a rede se obrigava a descobrir novos padrões para diferenciar as espécies, para testar as épocas, essa versão adotou 40 épocas, a fim de ver se o prolongamento da rede seria um problema.

Por fim, através das análises de todas as bases deu para concluir que o melhor modelo para seu utilizar em uma análise de espécies seria uma rede neural com a camada densa final, três camadas convolucionais, nós nos tamanhos de 32 ou 64, duas camadas de *pooling*, e três ativações, sendo duas *ReLU* e uma *Sigmoid*, atuando em 40 épocas. Essas duas redes apresentadas foram as que obtiveram o melhor resultado nas análises, e se mostram muito prestativas para efetuar essa ação.

Em geral, foi possível comprovar a grande capacidade que uma rede neural artificial pode possuir, e como suas camadas podem fazer diferença no resultado final. Essas redes se mostram muito úteis não somente na diferenciação de imagens, mas também em outras áreas, com segurança, saúde, educação, entre outras. Se mostra como sendo uma das grandes tecnologias a serem adotadas em um futuro próximo para a execução de tarefas no nosso dia a dia.

REFERÊNCIAS BIBLIOGRÁFICAS

[1] RUSSELL, Stuart; NORVIG, Peter. **Inteligência Artificial**. 3. ed. Rio de Janeiro: Editora LTC, 2013.

- [2] KINLAW, C. Ryan; DUNLAP, Linda L.; D'ANGELO, Jeffrey A. **Relations between faculty use of online academic resources and student class attendance**. 2012. Disponível em: <www.elsevier.com/locate/compedu>. Acesso em: 14 abr. 2019.
- [3] AZOFF, Michael. **Machine Learning in Business Use Cases**. Abril 2015. Disponível em:<http://images.nvidia.com/content/pdf/Ovum_Machine_Learning_in_Business_Use_Cases.pdf>. Acesso em: 24/06/2019.
- [4] SILVA, Ivan Nunes da; SPATTI, **Dailo Hernane**; FLAUZINO, Rogério Andrade. **Redes Neurais Artificiais: Para Engenharia e Ciências Aplicadas**. 2. ed. São Paulo: Artliber, 2016
- [5] MAZZA, Leonardo Oliveira. **APLICAÇÃO DE REDES NEURAIAS CONVOLUCIONAIS DENSAMENTE CONECTADAS NO PROCESSAMENTO DIGITAL DE IMAGENS PARA REMOCÃO DE RUÍDO GAUSSIANO**. TCC (Graduação) - Curso de Engenharia Controle e Automação, Universidade Federal do Rio de Janeiro, Rio de Janeiro, 2017
- [6] MOREIRA, Catarina. Neurónio. **Revista de Ciência Elementar**, Lisboa, v. 1, n. 1, p.1-3, out. 2013. Trimestral. -. Disponível em: <<https://rce.casadasciencias.org/rceapp/art/2013/008/>>. Acesso em: 27 out. 2019.
- [7] OLIVEIRA, Patrick Francisco. ANÁLISE DE DESEMPENHO DE UM ALGORITMO DESENVOLVIDO PARA SOLUÇÃO DE DEEP LEARNING UTILIZANDO REDES NEURAIAS CONVOLUCIONAIS PARA ANÁLISE DE CONTRASTE DE IMAGENS. **Revista Ubiquidade**, Jundiaí, v. 2, n. 1, p.84-105, jan. 2019. Semestral. Disponível em: <<http://www.portal.anchieta.br/revistas-e-livros/ubiquidade/pdf/2019/artigo-ubiquidade-vol2-artigo5.pdf>>. Acesso em: 4 nov. 2019.
- [8] LEVINE, Robert I.; DRANG, Diane E.; EDELSON, Barry. **INTELIGÊNCIA ARTIFICIAL E SISTEMAS ESPECIALISTAS: APLICAÇÕES E EXEMPLOS PRÁTICOS**. São Paulo: Mcgrowhill, 1988.
- [9] RICH, Elaine; KNIGH, Kevin. **Inteligência Artificial**. 2. ed. São Paulo: Mcgrowhill, 1995.
- [10] AGUIAR, Gabriel Jonas; ZARPELÃO, Bruno Bogaz; BARBON JUNIOR, Sylvio. **Melhoria de Contraste em Imagens Digitais baseado em Inteligência Artificial**. Londrina, 2017.
- [11] FERREIRA, Alessandro dos Santos. **Redes Neurais Convolucionais Profundas na Detecção de Plantas Daninhas em Lavoura de Soja**. 2017. 80 f. Dissertação (Mestrado) - Curso de Pós Graduação em Ciências da Computação, Universidade Federal de Mato Grosso do Sul, Campo Grande, 2017.
- [12] LIU, Danqing. **A Practical Guide to ReLU: Start using and understanding ReLU without BS or fancy equations**. 2017. Disponível em: <<https://medium.com/@danqing/a-practical-guide-to-relu-b83ca804f1f7>>. Acesso em: 23 nov. 2019.
- [13] PALMIERE, Sérgio Eduardo. **Introdução a Redes Neurais Artificiais**. 2016. Disponível em: <<https://www.embarcados.com.br/redes-neurais-artificiais-introducao/>>. Acesso em: 23 nov. 2019. (site das funções)
- [14] BRIXIUS, Nathan. **The Logit and Sigmoid Functions**. 2016. Disponível em: <<https://nathanbrixius.wordpress.com/2016/06/04/functions-i-have-known-logit-and-sigmoid/>>. Acesso em: 23 nov. 2019.
- [15] RIBEIRO, Amanda Gonçalves. **Função Linear**. Disponível em: <<https://mundoeducacao.bol.uol.com.br/matematica/funcao-linear.htm>>. Acesso em: 24 nov. 2019.
- [16] KINSLEY, Harrison. **How to use your trained model: Deep Learning basics with Python, TensorFlow and Keras** p.6. 2018. Disponível em:

<<https://pythonprogramming.net/using-trained-model-deep-learning-python-tensorflow-keras/>>. Acesso em: 24 nov. 2019.

[17] MICROSOFT. **Dogs vs Cats dataset from Microsoft**. Disponível em: <<https://www.microsoft.com/en-us/download/confirmation.aspx?id=54765>>. Acesso em: 24 nov. 2019.

[18] WANG, Pu et al. Rectified-Linear-Unit-Based Deep Learning for Biomedical Multi-label Data. **Interdisciplinary Sciences: Computational Life Sciences**, [s.l.], v. 9, n. 3, p.419-422, 11 nov. 2016. Springer Science and Business Media LLC. <http://dx.doi.org/10.1007/s12539-016-0196-1>

[19] SOARES, Vinicius. **“Nobel da Computação” vai para os pais do Deep Learning**. Disponível em: <<https://www.institutodeengenharia.org.br/site/2019/04/01/nobel-da-computacao%E2%80%8B-vai-para-os-pais-do-deep-learning/>>. Acesso em: 8 dez. 2019.

APLICAÇÃO E ANÁLISE DO MÉTODO ÁGIL SCRUM EM UM PROJETO DE “M.E.S [MANUFACTURING EXECUTION SYSTEM]”

USE AND ANALYSIS OF THE AGILE SCRUM METHOD IN A PROJECT OF “M.E.S [MANUFACTURING EXECUTION SYSTEM]”

Mário Vinícius VIEIRA

mariovinicius18@hotmail.com

MBA em Gestão de Projetos. PECEGE.

Cecilia SOSA ARIAS PEIXOTO

cecilia.sosaarias@gmail.com

Orientadora. PECEGE.

Resumo

Nos últimos anos, o mundo tem passado por diversas transformações e para acompanhar de forma eficaz essas mudanças, que ocorrem de maneira rápida, o gerenciamento de projetos tem sido aplicado, pois é focado nos objetivos e também nas prioridades da empresa. Neste contexto, este artigo apresenta a aplicação da metodologia de gestão de projetos, Scrum, em uma empresa multinacional do setor de tecnologia da informação. Essa empresa é prestadora de serviço e atua no ramo automobilístico onde existe grande demanda de projetos de “M.E.S [Manufacturing Execution System]” de curta duração e alto grau de criticidade. Estes projetos envolvem vários setores da fábrica e estão diretamente ligados ao ambiente de produção. Durante a aplicação da metodologia, dados foram obtidos para embasar a pesquisa apresentada neste trabalho. Foi elaborado um questionário, para todos os envolvidos no projeto, afim de identificar se o escopo foi cumprido, e se tudo o que foi proposto / requisitado foi atendido com sucesso, além de obter o feedback de todos quanto ao uso do método Scrum para os projetos.

Palavras-chave

gestão; agilidade; processo; escopo e sprint

Abstract

In recent years, the world has undergone several transformations. To effectively follow these changes, which occur quickly, project management has been applied, as it is focused on the objectives and also on the company's priorities. In this context, this article presents the application of the Scrum project management methodology in a multinational information technology company operating in the automotive industry, where there is a

great demand for “M.E.S [Manufacturing Execution System]” projects of short duration and high degree of criticality. These projects involve several sectors of the factory and are directly linked to the production environment. During the application of the methodology, data were obtained to support the research presented in this work. A questionnaire was prepared, for all those involved in the project, to identify if the scope was fulfilled, and if everything that was proposed / requested was successfully met, in addition to obtaining everyone's feedback on the use of the Scrum method for the projects.

Keywords

management, agility; process and sprint.

INTRODUÇÃO

Nos últimos anos, o mundo tem passado por diversas transformações, e para acompanhar de forma eficaz essas mudanças, que ocorrem de maneira rápida, o gerenciamento de projetos tem sido aplicado, pois é focado nos objetivos e também nas prioridades da empresa (VARGAS, 2016).

Sabbagh (2014), afirma que a metodologia tradicional gera processos complexos e volume de documentação e é um método que é focado em planejamento definido e detalhado, com informação de custo, escopo e cronograma, porém quando há a necessidade de alterações, o método se torna indesejado. Segundo Sutherland (2016), o método cascata, até o ano de 2005, era o método mais utilizado para desenvolver softwares, e era considerado devagar, inesperado e muitas vezes o resultado na entrega não era exatamente o que as pessoas estavam dispostas a pagar.

Em contraposição ao método tradicional surgiu a metodologia ágil ou leve. Ela foi proposta frente as dificuldades que as empresas tinham com os atrasos de projetos, complexidade tecnológica e também mudanças nas solicitações vindas do cliente. Ela é uma metodologia nova que utiliza uma comunicação mais informal, e que obtem poucas práticas, regras e volume de documentação. A utilização desse método na elaboração de projetos, é feita através da construção de reuniões presenciais e no fluxo de informações vindas do cliente (KERZNER, 2017).

De acordo com Prikladnicki et al. (2014), os métodos ágeis diferem dos métodos tradicionais em atender (foco), as pessoas e não nos processos, e isso auxilia na adequação as mudanças que ocorrem nos projetos e também no retorno rápido em resposta a oscilação do mercado.

Sabbagh (2014) afirma que existe interesse em adotar a agilidade em diferentes tipos de organizações e um dos métodos que incentiva essa agilidade é o Scrum. Sutherland (2016) confirma que criou o Scrum há vinte anos e complementa que ele e Ken Schwaber desenvolveram o Scrum como uma maneira mais confiável, suficiente e ágil no desenvolvimento dos softwares nas empresas de tecnologia.

Os métodos ágeis e o Scrum não necessariamente podem ser utilizados somente para o desenvolvimento de software, mas também podem ser aplicados a outros projetos (BROD, 2013).

De acordo com Audy (2015), a metodologia ágil Scrum tem como característica o efeito de fiscalização e de ajustes constantes. É considerada uma metodologia simples de aprender a técnica.

Devido aos benefícios da utilização da metodologia Scrum, foi identificado a possibilidade de aplicação do método com o intuito de contribuir com conhecimentos e informações. O principal objetivo deste trabalho, é apresentar os problemas existentes no método atual de gerenciamento dos projetos dessa empresa e aplicar o método ágil Scrum em um projeto de M.E.S da área, além de demonstrar se o escopo foi cumprido, e se tudo o que foi proposto / requisitado foi atendido com sucesso. Este trabalho também tem por objetivo obter o feedback dos participantes quanto a utilização do método Scrum para os projetos da área.

As informações presentes neste trabalho, podem gerar benefícios para a organização, já que esses projetos influenciam diretamente a satisfação dos clientes internos da organização.

Metodologia

De acordo com Marconi e Lakatos (2003) o método de pesquisa dialético existe as seguintes leis fundamentais: onde tudo se relaciona, tudo se transforma, existe a mudança qualitativa e luta dos contrários.

O trabalho tem como método de pesquisa dialético juntamente com uma pesquisa do tipo “survey”, que permitiu obter informações importantes sobre a aplicação do método Scrum (FREITAS et al., 2000).

Este trabalho foi dividido em algumas partes, sendo elas:

- Análise da Situação Problema.
- Planejamento da Implantação do Scrum.
- Implantação e acompanhamento do Scrum.
- Questionário.

Para identificar qual seria o projeto piloto ideal, foi utilizado o conselho de Mike Cohn, onde ele diz que nenhum projeto é adequado para ser o primeiro, no entanto ele destaca os principais pontos para a escolha: duração, tamanho, importância e engajamento do patrocinador (COHN, 2011).

Levando em consideração esses pontos, foi determinado a escolha do Projeto: Tracking System (Sistema de Rastreamento) onde o prazo médio deveria ser de aproximadamente de 2 a 3 meses, com 5 pessoas na equipe. É válido ressaltar que é um projeto de alta importância.

O questionário elaborado nesta pesquisa, foi dividido em duas partes, sendo a primeira parte com afirmações baseadas nos sete benefícios do método Scrum informados por Sabbagh (2014), a saber:

1. Entregas frequentes de retorno ao investimento dos clientes.
2. Redução dos riscos do projeto.
3. Maior qualidade no produto gerado.

4. Mudanças utilizadas como vantagem competitiva.
5. Visibilidade do progresso do projeto.
6. Redução do desperdício.
7. Aumento da produtividade.

O Scrum time respondeu o grau de concordância dessas afirmações. A escala utilizada foi de (LIKERT, 1932), e varia de Concordo Totalmente até Discordo Totalmente. A cada resposta é atribuído um número sendo 5 para Concordo Totalmente até 1 para Discordo Totalmente. A segunda parte do questionário, foi a descrição da principal vantagem e desvantagem que cada integrante do time identificou em alguns aspectos do Scrum.

Cenário

Apesar da empresa ser multinacional, faz pouco tempo que se estabilizou no Brasil (aproximadamente cinco anos), no início, priorizou sua instalação e adequação de processos de produção. Atualmente implementa soluções de melhorias em diversos setores da empresa. Até o momento nunca houve um método específico de trabalho para atendimento das demandas dos projetos, inclusive de software. Os projetos sempre tiveram métodos informais e dificuldades na determinação de prazo e custo dos projetos. Os principais projetos eram tratados por equipes internacionais que na maioria das vezes, solicitavam apenas ajudas operacionais, finalizavam o projeto e voltavam para a filial.

Nos últimos tempos os analistas têm encontrado muitos problemas no controle desses projetos, já que cada analista na maioria das vezes controla seus projetos de forma particular e com pouco envolvimento com a equipe. A demanda de projetos vêm aumentando e a quantidade de pessoas na equipe permanece a mesma o que contribui para uma má gestão dos projetos.

A equipe de atendimento dos projetos é composta por 3 analistas que possuem conhecimento no processo produtivo e prestam serviços nas diversas áreas de fabricação (Estrutura, Pintura e Montagem) de automóveis e experiência em desenvolvimento de “software”, 1 líder de desenvolvimento / integração de equipamentos e 1 gerente que auxilia nas questões que necessitam de reporte para a filial.

O departamento de PCP [Planejamento e Controle da Produção] estava com muitas dificuldades para rastrear em qual local da etapa de pintura o veículo se encontrava, visto que o dispositivo atual “PLC [Programmable Logic Controller – Controlador de Lógico Programável]” apesar de ser eficiente em controlar várias instalações na empresa, não demonstrava essas informações e causava retrabalhos e atrasos. A área de M.E.S foi requisitada afim de identificar / desenvolver melhorias no sistema de rastreamento de veículos, e assim surgiu a oportunidade de aplicação do Scrum.

Scrum

Para alcançar os objetivos, primeiramente foi necessário entender alguns termos que esse método sugere, como algumas terminologias e seus significados de acordo com Sabbagh (2014) e também Schwaber (2004), são:

- “Scrum Team” – Equipe que obtém informações e conhecimento necessário para alcançar as metas acordadas pelo “Product Owner”.
- “Product Owner” – Responsável por tomar conta do “Product Backlog”, ele desenvolve, altera, reordena e torna visível, quando achar necessário.
- “Scrum Master” – Responsável por não deixar que os impedimentos do trabalho, atrapalhem no alcance dos objetivos, incentiva na auto-organização portanto é considerado um facilitador.
- “Product Backlog” – É uma listagem de tudo que irá ser desenvolvido pela equipe com detalhes chaves, e com prioridades no alto. Pode conter objetivo dos clientes, melhorias, correções, informações técnicas, pesquisas, etc.
- “Sprints” – São ciclos do projeto, andamento do trabalho que será efetuado conhecido também como arrancadas, e devem ser feitos de forma rápida.
- “Sprint Backlog” – Plano de como será cumprido a listagem do “Product Backlog”.
- “Time Box” – Prazo pré-determinado para os eventos do Scrum.
- “Sprint Planning” – Reunião de trabalho para planejar o Sprint, é negociado o que será desenvolvido para alcançar a meta.
- “Meta do Sprint” – Objetivo a ser alcançado pelo “Scrum team”.
- “Daily Scrum” – Reunião rápida diária para responder as seguintes questões: O que você tem feito desde a última “Daily Scrum”? O que você planeja fazer até a próxima reunião? O que impede de seguir neste Sprint?
- “Release Burndown” – Demonstra a quantidade de trabalho que ainda falta ser executado.
- “Sprint Retrospective” – Última reunião que é feita antes do término do Sprint, é utilizada para realizar a inspeção do trabalho até o próximo Sprint.
- “Sprint Review” – Se refere ao feedback do cliente e das pessoas do time, isso serve como material para modificações no “Product Backlog”.
- O Scrum, é dividido em algumas fases, começando com a preparação do “Product Backlog” e “Sprint Backlog”, os “Sprints” acontecem de 1 a quatro semanas, contemplando os “Daily Scrum”, e são seguidas de uma entrega funcional (Brod, 2013).

Sabbagh (2014) diz que para ter sucesso nos ciclos do Scrum, as “Sprints” deve-se seguir a seguinte ordem: “Sprint Planning” – Desenvolvimento – “Sprint Review” e Sprint Retrospective.

Após o término do projeto, foi aplicado um questionário ao Scrum time (Apêndice 1), que consta de uma avaliação da aplicação do método Scrum neste projeto para posterior análise.

Resultados e discussão

Análise da Situação Problema

Durante a fase de Análise da Situação Problema foi realizado um levantamento de informações com a equipe de atendimento dos projetos de T.I do M.E.S, com os principais problemas atuais em relação aos projetos executados:

- Ausência de um método específico de gerenciamento de projeto.
- Ausência de um responsável direto e indireto pelo projeto.
- Ausência da reunião de abertura e escopo dos projetos.
- Informações das necessidades dos projetos vêm de maneira informal, com urgência e ruídos de informação (exemplo: dificuldade de comunicação devido ao idioma).
 - Necessidade de apoio de outras áreas as quais não tem conhecimento que existe um projeto em andamento.
 - Atraso de desenvolvimento, atraso em compras quando existe a necessidade.
 - Descumprimento no prazo: 100% dos projetos não são cumpridos no prazo estipulado.
 - Mudanças de escopo: durante o andamento do projeto existem muitas alterações de escopo.
 - Nível de importância do projeto é alterado ao longo do desenvolvimento, deixando os envolvidos desinteressados e focando em outras prioridades ou projetos.
 - Áreas envolvidas não tem entrosamento e engajamento devido não conhecer as prioridades e as pessoas envolvidas.

O método tradicional de gerenciamento de projetos foi aplicado na empresa em outros momentos sem êxito. Entre os principais problemas se encontra a falta de cumprimento do cronograma e as entregas não aceitas pelo cliente (quando era feita alguma entrega ao cliente, este estava em desacordo com o produto entregue e gerava muitos retrabalhos para o time).

Planejamento da Implantação do Scrum

Para entender melhor o projeto, foi necessário envolver todos os interessados através de uma reunião na qual foi definido o escopo e o “Scrum team” para este projeto:

- “Product Owner” – Analista de sistemas do M.E.S (autor desse trabalho)
- “Scrum Master” – Coordenador do M.E.S
- Desenvolvedor 1 – Analista de sistemas do M.E.S
- Desenvolvedor 2 – Analista de sistemas do M.E.S
- Desenvolvedor 3 – Analista de manutenção
- Desenvolvedor 4 – Analista de manutenção

O autor desse projeto assumiu o papel de Product Owner visto que o requisitante já estava envolvido com outros trabalhos. O coordenador do M.ES foi o Scrum Master, pois tem conhecimento geral da área e facilidade na remoção de impedimentos. Os desenvolvedores 1, 2,3 e 4 possuíam habilidades em desenvolvimentos de aplicações Web, conhecimento operacional e de programação de PLC, e foram pessoas técnicas importantes na execução das atividades as quais eram necessárias para o cumprimento e bom andamento do projeto. A tabela 1 apresenta o escopo do projeto.

Tabela 1. Escopo do Projeto

<p>Justificativa: Devido ao amplo tamanho do parque de produção que possui dois pisos, sendo o térreo com onze linhas de produção e o primeiro andar com quatorze linhas, o projeto surgiu devido à necessidade de se localizar uma determinada carroceria com urgência para fins de auditoria. Além da necessidade de rastreabilidade das carrocerias, também foi levantado um ponto referente às principais especificações de cada carroceria para melhor controle e acompanhamento da produção.</p>
<p>Gerente do Projeto: M.E.S: Mário Vinicius</p>
<p>Partes interessadas Controle de Produção: Bruno / André (Patrocinadores) Controle de Qualidade: Amanda / Alan Manutenção: Ederson / Carlos M.E.S: Anderson (Gerente) Evandro / Mário Vinicius</p>
<p>Premissas: Time de manutenção providenciar as informações nas Db's do PLC. Time de manutenção providenciar o mapeamento das linhas para elaborações dos layouts.</p>
<p>Restrições: Considera-se que toda infraestrutura para hospedagem da aplicação já se encontra em operação bastando apenas publica-la. O time de manutenção se responsabiliza em obter as informações do PLC, trata-las e escreve-las nas respectivas DB's de alcance do M.E.S.</p>
<p>Descrição do escopo: Será desenvolvido um sistema web que utiliza a tecnologia "silverlight" juntamente com a linguagem de programação C#. Aplicação deve ser acessível de qualquer estação de trabalho da fábrica através da web (IE 11) A aplicação irá possuir uma tela principal com todas as linhas de produção mostrando a quantidade total de carros por linha e status de cada linha. Ao clicar em uma determinada linha, uma nova janela se abre mostrando a linha em detalhe com as carrocerias e as especificações. Após análise dos dados a tela pode ser fechada e retorna-se para a tela principal.</p>
<p>Exclusões: Não há</p>
<p>Risco: Demora na obtenção das informações pelo time de manutenção. Limitação de tempo devido outras atividades dos envolvidos (mudança de prioridades). Configurações erradas no software de comunicação (TM) que possam afetar o sistema de informações da produção.</p>
<p>Entregas: Sistema em operação. Manual de instruções. Treinamento para usuários chave.</p>

Fonte: Resultados originais da pesquisa

A tabela 1 mostra a definição do escopo do projeto, onde se tem especificado a justificativa do projeto, partes interessadas, premissas, restrições, exclusões, riscos e entregas.

Após a definição do escopo e time do projeto, foi realizado o treinamento do "framework" Scrum com os envolvidos, afim de apresentar a ferramenta e preparar aqueles que não tinham conhecimento.

Implantação e acompanhamento do Scrum

Para realizar a implantação do Scrum, na reunião de "Product Backlog", o "Product Owner" e a equipe montou os "Backlogs" com as principais tarefas e funcionalidades a serem executadas, considerando o nível de prioridade determinado pelo "Product Owner". A equipe encontrou dificuldades de separar as tarefas, visto que se tratava da

primeira vez. Apesar do “Product Backlog” conter somente 5 Backlog Item, cada item possui tarefas complexas. A Tabela 2, apresenta o “Product Backlog” do projeto.

Tabela 2. “Product Backlog”

“PRODUCT BACKLOG”	
Alta prioridade	Backlog Item
	1 Mapeamento das linhas / estações
	2 Configuração dos blocos de memória do PLC
	3 Criação e configuração das TAG's de Interface no software TM connector
	4 Desenvolvimento do layout e componentes
	5 Configuração das Tag's nos componentes e programação de funcionalidades.
Baixa prioridade	

Fonte: Resultados originais da pesquisa

A tabela 2 mostra a listagem de todas as tarefas que precisam ser realizadas para se entregar um incremento do produto, as tarefas são organizadas de alta prioridade no topo até a baixa prioridade.

O “Scrum Master” informou que as reuniões de “Daily Scrum”, teriam duração de no máximo 15 minutos, e este ficou auxiliando a equipe no acompanhamento do projeto.

Os “Sprints Planning” foram realizados antes do início das “Sprints”, e foram reuniões de planejamento dos “Sprints”, como este projeto tem praticamente as mesmas tarefas para diversas linhas presentes na fase de pintura do veículo, existiu facilidade em definir os “Sprints” Backlog, e o planejamento de quantos “Product Backlog” seriam atendidos dentro do próximo “Sprint”. A Tabela 3 apresenta o “Sprint Backlog”.

Tabela 3. “Sprint Backlog”

“SPRINT BACKLOG”				
SPRINT 1				
Linhas	Atividade	Responsável	Status	Entrega do Produto
Overvi	Elaboração de um esboço da linha com as estações	Dev 3 / Dev 4	Realizado	Status de cada linha: Rodando:

	Elaboração de um documento com todas as estações das linhas com os respectivos endereços de PLC	Dev 3 / Dev 4	Realizado	Verde, Parada: Amarelo, Alarme: Vermelho
	Elaboração de documento contendo a linha e os nomes das TAG's para cada estação	Dev 1 / Dev 2	Realizado	Quantidade de carrocerias por linha
	Utilização do esboço da linha para desenhar no software de desenvolvimento	Dev 1 / Dev 2	Realizado	Ao clicar em cada linha uma nova janela abre mostrando as informações das carrocerias
	Utilização do software Blend e linguagem C# para programação	Dev 1 / Dev 2	Realizado	Campo de busca pelo numero da carroceria
WBS	Elaboração de um esboço da linha com as estações	Dev 3 / Dev 4	Realizado	Status da linha
	Elaboração de um documento com todas as estações das linhas com os respectivos endereços de PLC	Dev 3 / Dev 4	Realizado	Quantidade de carrocerias na linha
	Elaboração de documento contendo a linha e os nomes das TAG's para cada estação	Dev 1 / Dev 2	Realizado	Exibir a linha com as respectivas carrocerias
	Utilização do esboço da linha para desenhar no software de desenvolvimento	Dev 1 / Dev 2	Realizado	Cada tipo diferente de carroceria deve ser exibido de uma cor diferente
	Utilização do software Blend e linguagem C# para programação	Dev 1 / Dev 2	Realizado	Exibir as especificações de cada carroceria quando colocar o mouse sobre a estação
P/T	Elaboração de um esboço da linha com as estações	Dev 3 / Dev 4	Realizado	Status da linha
	Elaboração de um documento com todas as estações das linhas com os respectivos endereços de PLC	Dev 3 / Dev 4	Realizado	Quantidade de carrocerias na linha
	Elaboração de documento contendo a linha e os nomes das TAG's para cada estação	Dev 1 / Dev 2	Realizado	Exibir a linha com as respectivas carrocerias
	Utilização do esboço da linha para desenhar no software de desenvolvimento	Dev 1 / Dev 2	Realizado	Cada tipo diferente de carroceria deve ser exibido de uma cor diferente
				Exibir as especificações de cada carroceria

	Utilização do software Blend e linguagem C# para programação	Dev 1 / Dev 2	Realizado	quando colocar o mouse sobre a estação
E/D Coat	Elaboração de um esboço da linha com as estações	Dev 3 / Dev 4	Realizado	Status da linha Quantidade de carrocerias na linha
	Elaboração de um documento com todas as estações das linhas com os respectivos endereços de PLC	Dev 3 / Dev 4	Realizado	Exibir a linha com as respectivas carrocerias
	Elaboração de documento contendo a linha e os nomes das TAG's para cada estação	Dev 1 / Dev 2	Realizado	Cada tipo diferente de carroceria deve ser exibido de uma cor diferente
	Utilização do esboço da linha para desenhar no software de desenvolvimento	Dev 1 / Dev 2	Realizado	Exibir as especificações de cada carroceria
	Utilização do software Blend e linguagem C# para programação	Dev 1 / Dev 2	Realizado	quando colocar o mouse sobre a estação
ED Buffer	Elaboração de um esboço da linha com as estações	Dev 3 / Dev 4	Realizado	Status da linha Quantidade de carrocerias na linha
	Elaboração de um documento com todas as estações das linhas com os respectivos endereços de PLC	Dev 3 / Dev 4	Realizado	Exibir a linha com as respectivas carrocerias
	Elaboração de documento contendo a linha e os nomes das TAG's para cada estação	Dev 1 / Dev 2	Realizado	Cada tipo diferente de carroceria deve ser exibido de uma cor diferente
	Utilização do esboço da linha para desenhar no software de desenvolvimento	Dev 1 / Dev 2	Realizado	Exibir as especificações de cada carroceria
	Utilização do software Blend e linguagem C# para programação	Dev 1 / Dev 2	Realizado	quando colocar o mouse sobre a estação

Fonte: Resultados originais da pesquisa

A cada duas semanas foram realizados os “Sprints” (“Time Box” estipulado), e a cada final dos “Sprints”, era entregue algum incremento do produto, como: elaboração de documentos, desenho do software e programação, ou seja, parte do sistema web realizado. Nesse período houveram algumas modificações por parte dos clientes internos, e causou um pouco de frustração no time, no entanto, as modificações realmente eram necessárias.

No final dos “Sprints”, também eram realizados os “Sprints Review”, que auxiliavam na necessidade de ajuste ou adaptabilidade do que estava sendo realizado, e também os “Sprints Retrospective”, que guiava na identificação do que foi feito de errado, e poderia ser melhorado nas próximas “Sprints”.

Durante os “Sprints” foram identificadas algumas mudanças de prioridade do “Product Backlog” e foram incluídas pelo “Product Owner”.

O projeto foi acompanhado através do “Burndown Chart”, que auxiliou no monitoramento das tarefas, e também na identificação dos atrasos das entregas. A Figura 1 apresenta o gráfico de “Burndown”.

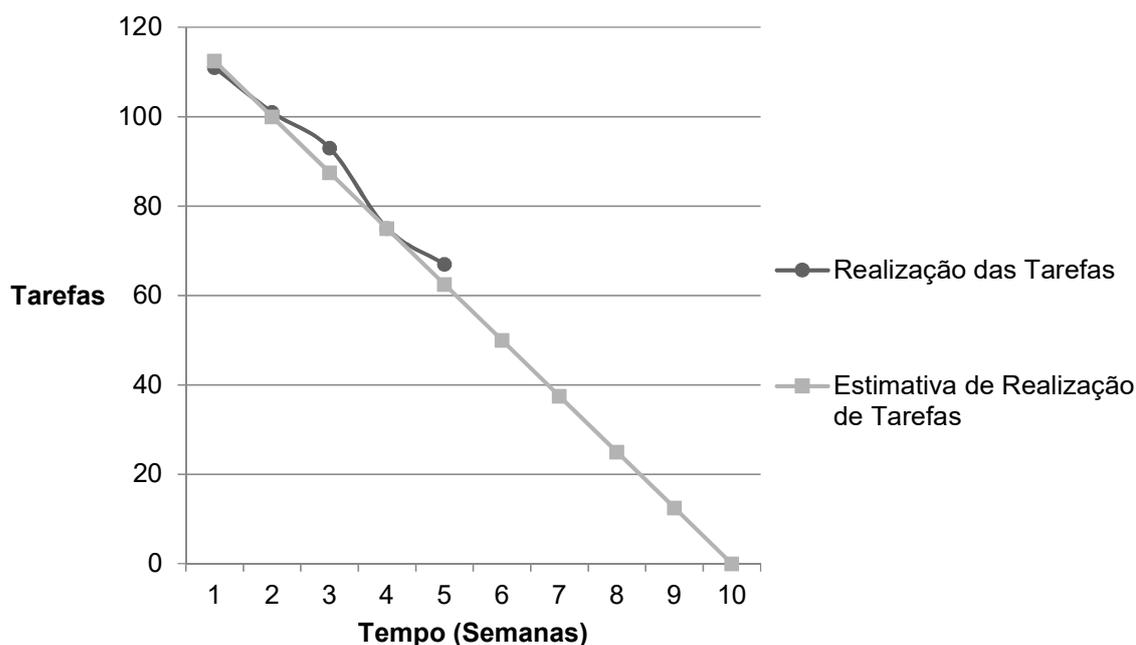


Figura 1. Gráfico de “Burndown”

Fonte: Resultados originais da pesquisa

Ao longo do projeto, identificou-se que não houve muito desvio do planejado. Foram necessárias 5 “Sprints” para a realização do projeto. O tempo total do projeto foi em torno de 3 meses para alcançar o término e não houveram muitas variações e atrasos no cronograma.

O cliente ficou satisfeito com a entrega e sugeriu aplicar o mesmo projeto para outras linhas de produção. Foram necessários alguns manuais com as instruções e procedimentos para auxiliar na utilização da ferramenta web.

Nas primeiras reuniões de “Sprints”, o autor deste trabalho questionou se houveram alguns benefícios na utilização do Scrum e foi feito um levantamento das principais melhorias identificadas:

- Reunião de abertura e escopo gerou informações necessárias de priorização.

- O time afirmou que o “Product Backlog” é uma parte chave para o Scrum, já que ele é base e guia para as “Sprints”.
- A motivação do time aumentou, visto que cada um definiu o que iria entregar até o final das “Sprints”. O membro não se sentia sobrecarregado, porque ele junto a equipe determinava o que seria entregue.
- As reuniões diárias facilitaram na remoção de dúvidas que surgiam ao decorrer da Sprint.
- Quando havia dificuldade na execução de alguma tarefa, a equipe se mobilizava para resolver o problema em equipe, com a ajuda do “Scrum Master”.
- O “Sprint Backlog” foi o que avançou e atrasou as “Sprints”, dessa forma, todos concordaram que a dedicação nesse ponto é muito importante.
- Projeto via Scrum passou a ser mais valorizado na empresa, pois houve uma formalização do projeto junto aos envolvidos e também a diretoria.

A Tabela 4 apresenta os resultados da primeira parte do questionário de pesquisa que contém as afirmações baseadas nos sete benefícios do método Scrum informados por Sabbagh (2014).

Tabela 4. 1º Resultado

Afirmações	Nível de concordância				Discordo Totalmente
	Concordo Totalmente	Concordo	Indeciso	Discordo	
1. Entregas frequentes de retorno ao investimento dos clientes	3	3	0	0	0
2. Redução dos riscos do projeto	2	3	1	0	0
3. Maior qualidade no produto gerado	4	2	0	0	0
4. Mudanças utilizadas como vantagem competitiva	0	4	2	0	0
5. Visibilidade do progresso do projeto	5	1	0	0	0
6. Redução do desperdício	1	5	0	0	0
7. Aumento da produtividade	4	2	0	0	0
Total	45%	48%	7%		

Fonte: Resultados originais da pesquisa

Os resultados demonstram que a equipe em sua maioria concorda com os benefícios causados pelo método Scrum os quais Sabbagh (2014) afirma na literatura. Somente 7 % do total de respostas informou estar indeciso em relação a essas afirmações.

Acredita-se que esse resultado se deva ao fato de que as entregas aos finais das “Sprints” colaboraram na obtenção do feedback do cliente ao longo do projeto, reduzindo a possibilidade de falhas e riscos, já que todos os integrantes estavam focados nas atividades e o cliente envolvido no andamento. Mesmo havendo modificações ao longo do projeto, isso aproximava ao que o cliente realmente precisava. A visibilidade das atividades trouxe confiança além de possibilitar entregas com qualidade.

Apesar de no passado não ser aplicado por completo o método tradicional de gerenciamento de projetos, demorava-se muito tempo, porém após a aplicação do método Scrum a equipe se manteve envolvida e focada e não houve muito espaço para obter desperdício, principalmente relacionado ao tempo, já que o projeto foi tratado como formal e a equipe sentia que o resultado dos trabalhos eram visíveis, pois todos se empenhavam em entregar as tarefas no prazo.

Com o auxílio das reuniões diárias, o time conseguiu prever possíveis problemas e/ou impedimentos antes mesmo deles acontecerem, e isso auxiliava a manter o projeto dentro do cronograma. A produtividade aumentou, pois equipe se manteve entrosada, com comunicação e colaboração constante de todos, isso gerou o cumprimento do escopo, prazo e requisitos do cliente.

Em questionamento ao atendimento do projeto, o cliente informou estar satisfeito com as entregas do projeto. Na segunda parte do questionário, demonstrado a partir da Tabela 5, foi possível obter o feedback do time quanto ao método Scrum.

Tabela 5. 2º Resultado

Terminologia	Integrante	Vantagem	Desvantagem
“Scrum Team”	1	Passa mais confiança	Levar pessoas nas costas
	2	Divisão das responsabilidades	Alguns integrantes se esforçam mais que outros
	3	Colaboração entre integrantes	Manter o time todo em foco
	4	Trabalho em equipe	Algumas pessoas acreditam ser melhores que outras
	5	Todos engajados em um propósito	Abrir mão das vontades próprias
	6	Conhecimento técnico elevado	Precisa saber um pouco de tudo
“Product Owner”	1	Existe um responsável pelo produto	Falta de experiência
	2	Define o que precisa ser entregue	-
	3	Ordena o que é prioridade	Nada
	4	Definição do produto	Realiza muitas alterações
	5	Valida e aceita as entregas	Acelera muito o time
	6	É dono do produto	Acredita que tudo é muito fácil
“Scrum Master”	1	Foco no objetivo	Nenhuma
	2	Ajuda na resolução de dificuldades	Muita cobrança
	3	Resolução de problemas	Nada
	4	Protege a equipe	Acredita ser o dono da equipe
	5	Ajuda no gerenciamento	Nenhuma
	6	Facilita o trabalho	Não vejo desvantagens
“Product Backlog”	1	Melhora no planejamento	Nenhuma
	2	Facilidade	-
	3	Visão geral	Nada
	4	Saber as dificuldades futuras	Muitas alterações
	5	Prioriza as tarefas	Nenhuma
	6	Nível de prioridade de entregas	Muda-se muito

“Sprints”	1	Soluções rápidas	Correria para preparar entregas a tempo
	2	Entregar uma parcela utilizavel do produto	Muito curta para determinadas entregas.
	3	Observar o projeto saindo do papel	Nada
	4	Definição do tempo para se entregar determinado parte utilizavel do produto	Curtas
	5	Divisão de um projeto em ciclos	Muitas “Sprints”
	6	Entregar uma parte funcional do produto	Não vejo desvantagens
“Sprint Backlog”	1	Agilidade na execução das tarefas	Perde-se algum tempo para definição antecipada
	2	Existencia de um plano pre-determinado	-
	3	Saber como deve se agir em determinadas tarefas	Nada
	4	Diminuição de erros	-
	5	Mostra o caminho	Nenhuma
	6	Como se proceder em cada etapa	Não vejo desvantagens
Time Box	1	Diminui perda de tempo	As vezes tempo ficou apertado
	2	Melhor controle doTempo	-
	3	Auto controle	Tarefas com tempo curto
	4	Facilita as entregas	-
	5	Aumento do foco	Tempo curto
	6	Tempo determinado para entregar	Não vejo desvantagens
“Sprint Planning”	1	Saber como alcançar a meta	Tomar decisões precipitadas
	2	Acerto sobre o que cada membro precisa fazer	-
	3	Definição do andamento da sprint	Algumas vezes necessita ajuste
	4	Detalhamento de como executar a sprint	-

Meta do Sprint	5	Define-se como o time irá trabalhar para finalizar a sprint	Nenhuma
	6	Diferentes opiniões	Não vejo desvantagens
	1	Desafio de entrega	Conflito na definição por parte do time e “Product Owner”
	2	Definição clara do entregável	-
	3	Saber onde precisa chegar	Nada
	4	Determinar objetivo	-
“Daily Scrum”	5	Entender o que precisa ser entregue	Nenhuma
	6	Define o objetivo	Não vejo desvantagens
	1	Verifica-se o bom andamento das tarefas	Nem todos comparecem
	2	Melhora na qualidade	-
	3	Ajuda na comunicação	Nada
	4	Identifica-se problemas	-
Release “Burndown”	5	Monitora o andamento	Nenhuma
	6	Expor necessidades	Não vejo desvantagens
	1	Gestão Visual	Nenhuma
	2	Verifica-se o andamento	-
	3	Identifica do ritmo do time	Nada
	4	Bom para identificar atraso	-
Sprint Retrospective	5	Podemos verificar a velocidade do time	Nenhuma
	6	Gestão a vista	Não ajuda em nada o time
	1	Alinhamento dos trabalhos	Nenhuma
	2	Realizar ajustes finos	-
	3	Inspecionar o trabalho	Não é necessário
	4	Verifica se tudo está OK	-
	5	Analisa se os trabalhos estão corretos	Às vezes atrapalha
	6	Verifica se existe inconsistência na entrega	Não vejo desvantagens

	1	Melhoria continua	Nenhuma
	2	Validação das entregas	-
“Sprint Review”	3	Receber o feedback	Nada
	4	Identificar problemas e corrigir	-
	5	Definições entre cliente e time	Gera muitas discussões
	6	Checagem e validação das entregas	Mudança de escopo

Fonte: Resultados originais da pesquisa

Alguns resultados estavam repetidos e nem todos os campos foram preenchidos, porém ainda assim foi possível verificar que, algumas vantagens e facilidades identificadas também são encontradas nas afirmações de Cohn (2011), como: maior satisfação e engajamento no trabalho pelos funcionários, qualidade e visibilidade (transparência).

Parte do objetivo da pesquisa foi apresentar o feedback de todos quanto ao método Scrum para os projetos de M.E.S e foi possível identificar já nos primeiros “Sprints” que a equipe sentiu alguns impactos positivos, como motivação, agilidade no andamento do projeto, valorização do trabalho e colaboração da maior parte dos envolvidos.

Algumas informações importantes em relação a aplicação do Scrum que foram identificadas na pesquisa são:

- Importância da escolha de um bom Project Owner, pois este deve ter um visão clara do que o cliente (requisitante) deseja.
- Obter um “Product Backlog” completo e detalhado.
- O escopo precisa ser bem definido. Essas informações são de extrema importância para poder realizar as entregas durante as “Sprints”.

Mesmo tendo ocorrido dificuldades ao longo do planejamento e monitoramento do trabalho, muitas delas devido a falta de experiência da equipe frente ao método Scrum o qual nunca foi aplicado na área, pode-se dizer que os resultados obtidos ao final da implantação foram satisfatórios e positivos em frente a situação problema em que a área se encontrava.

Nesta pesquisa os questionários foram focados apenas ao Scrum Time e não abordou questões direcionadas ao cliente, e devido a este fato, seguem três sugestões que tratam o feedback do cliente referente a aplicação do Scrum para pesquisas futuras:

- O projeto atendeu aos requisitos solicitados? Comente sobre os resultados atingidos em comparação aos resultados esperados.
- Como você avalia o desempenho da equipe em frente ao cronograma e fases do projeto?
- Qual aspecto do andamento ou entrega do projeto que precisa ser melhorado?

CONCLUSÕES

É possível concluir que o trabalho alcançou o objetivo de aplicar e analisar a implementação do método ágil Scrum em um projeto de M.E.S da área. Em relação ao projeto desenvolvido foi possível cumprir o escopo e atender tudo o que foi proposto com sucesso. Além de também obter o feedback de todos quanto ao método do Scrum para o projeto da área.

Nesse projeto todos os esforços do time envolvido, estavam sobre esse projeto e método, e é natural e evidente que a equipe deu prioridade e dedicação a esse projeto em frente aos demais da área.

Apesar de no passado não ser utilizado o método tradicional em sua maneira plena, é possível acrescentar que o trabalho com o método Scrum gerou algumas lições aprendidas, dentre elas a redução no tempo no desenvolvimento de um projeto (burocracia) e identificação de maneiras mais eficientes (método ágil Scrum) para gerenciar novos projetos. O método mostrou que contempla a qualidade do produto final e também o engajamento e satisfação da equipe.

Mesmo as reuniões sendo estipuladas para ocorrer todos os dias, houve falha na presença de alguns integrantes, e apesar da ausência a equipe não encontrou tanta dificuldade de planejamento, e ao longo das “Sprints”, se sentiram mais confortáveis.

Dentre alguns resultados, foi possível destacar alguns sinalizadores, que indicaram aspectos a serem revistos como: tempo das “Sprints”, não comparecimento de todos no “Daily Scrum”, necessidade do “Release Burndown” e forma de lidar com o “Sprint Retrospective” e “Sprint Review”. Esse são pontos que podem ser melhorados em futuros projetos. A falta de experiência da equipe pode ter causado esses sinalizadores, dentre outros, portanto é necessário aplicar o Scrum em novos projetos para realizar uma comparação.

No contexto geral o projeto foi bem-sucedido, com um cronograma atendido dentro prazo. Os recursos foram explorados dentro do possível e apesar das limitações, o trabalho traz importantes informações para empresa e equipe envolvida.

A conclusão, portanto, foi que a pesquisa contribuiu para a empresa e o trabalho se mostrou eficiente, e gerou conhecimento para a equipe envolvida, possibilitando a aplicação desse novo método para gerenciar os novos projetos da área.

REFERÊNCIAS BIBLIOGRÁFICAS

AUDY, J. *Scrum 360: um guia completo de prático de agilidade*. Editora Casa do Código. Disponível em: < https://books.google.com.br/books?hl=pt-BR&lr=&id=imaCCwAAQBAJ&oi=fnd&pg=PT5&dq=CRISTAL%3B+WILDT%3B+PRIKLADNICKI,+2008+livro&ots=-S5C-rwQP8&sig=SBQEuo-Q_5cgrW3WbLbGta11N8#v=onepage&q&f=false>. Acesso em: 26 abr.2018.

BROD, C. *Scrum: guia prático para projetos ágeis*. Editora Novatec, São Paulo, SP, Brasil. 2013.

COHN, M. *Desenvolvimento de software com Scrum: aplicando métodos ágeis com sucesso*. Editora Artmed, Porto Alegre, RS, Brasil. 2011.

FREITAS, H.; OLIVEIRA, M.; SACCOL, A. Z.; MOSCAROLA, J. *O método de pesquisa survey*. Revista de Administração 35: 105-12. 2000.

KERZNER, H. *Gestão de projetos: as melhores práticas*. Editora Bookman, Porto Alegre, RS, Brasil. 2017.

LIKERT, R. A technique for the measurement of attitudes. *Archives of Psychology* 140: 1-55. 1932.

MARCONI, M.; LAKATOS, E. *Fundamentos de metodologia científica*. Editora Atlas, São Paulo, SP, Brasil. 2003

PRIKLADNICKI, R.; WILLI R.; MILANI F. *Métodos ágeis para desenvolvimento de software*. Editora Bookman, Porto Alegre, RS, Brasil. 2014.

SABBAGH, R. *Scrum: gestão ágil para projetos de sucesso*. Editora Casa do Código. 2014. Disponível em: < https://books.google.com.br/books?hl=pt-BR&lr=&id=pG-CCwAAQBAJ&oi=fnd&pg=PT9&dq=gest%C3%A3o+de+projetos+%C3%A1gil&ots=ESMyvRECw9&sig=EZd1NB1_DKjTn5MOLWYtmdhZzmc#v=onepage&q=gest%C3%A3o%20de%20projetos%20%C3%A1gil&f=false>. Acesso em: 26 abr.2018.

SCHWABER, K. *Agile project management with Scrum*. Editora Microsoft Press. 2004. Disponível em: <<https://books.google.com.br/books?hl=pt-BR&lr=&id=6pZCAwAAQBAJ&oi=fnd&pg=PT9&dq=scrum&ots=kbpURX7tfU&sig=v6R2t9rsh4VKUPxHhJ6ilWt7YgY#v=onepage&q=scrum&f=false>>. Acesso em: 26 abr.2018.

SUTHERLAND, J. *A arte de fazer o dobro de trabalho na metade do tempo*. Editora Leya, São Paulo, SP, Brasil. 2016.

VARGAS, R. *Gerenciamento de projetos: estabelecendo diferenciais competitivos*. Editora Brasport, Rio de Janeiro, RJ, Brasil. 2016.

APÊNDICE

Apêndice 1. Avaliação do método ágil Scrum aplicado em projeto de M.E.S.

Cargo:

Benefícios avaliados:

1. Entregas frequentes de retorno ao investimento dos clientes

() Concordo Totalmente

() Concordo

() Indeciso

() Discordo

() Discorda Totalmente

2. Redução dos riscos do projeto

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

3. Maior qualidade no produto gerado

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

4. Mudanças utilizadas como vantagem competitiva

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

5. Visibilidade do progresso do projeto

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

6. Redução do desperdício

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

7. Aumento da produtividade

- Concordo Totalmente
- Concordo
- Indeciso
- Discordo
- Discorda Totalmente

Descreva a principal vantagem e desvantagem que identificou nos seguintes aspectos do Scrum:

	Vantagem	Desvantagem
“Scrum Team”		
“Product Owner”		
“Scrum Master”		
“Product Backlog”		
“Sprints”		
“Sprint Backlog”		
“Time Box”		
“Sprint Planning”		
Meta do “Sprint”		
“Daily Scrum”		
“Release Burndown”		
“Sprint Retrospective”		
“Sprint Review”		

IMPLEMENTANDO UMA PLATAFORMA BIG DATA PARA VISUALIZAÇÃO DE DADOS GERADOS POR DISPOSITIVO IOT

IMPLEMENTING A BIG DATA PLATFORM FOR VIEWING DATA GENERATED BY IOT DEVICE

Jair Campos de Oliveira JUNIOR

jair.ntj@hotmail.com

Pós-Graduação Lato Sensu em BI com Big Data,

Centro Universitário Padre Anchieta, Jundiaí/SP

Juliano SCHIMIGUEL (Orientador)

schimiguel@gmail.com

Coordenação Pós-Graduação Lato Sensu em BI com Big Data,

Centro Universitário Padre Anchieta, Jundiaí/SP

Resumo

O volume de dados vem crescendo a cada dia e são gerados de diferentes fontes, como redes sociais, mídia, Data Warehouse, sensores e de dispositivos *IoT*, tanto da indústria como de dispositivos portáteis. Diferentes plataformas de serviços *Cloud* vêm desenvolvendo métodos e ferramentas para extrair e transformar estes dados em algo de valor. A internet das coisas está em todos os setores e mercados, estando presente no setor automobilístico, na saúde, nas casas, em fabricas, em produtos, na segurança pública e privada, no meio ambiente, na agricultura e nos alimentos, e em outros seguimentos. Nesse artigo será desenvolvido um dispositivo *IoT* para coleta de dados gerados por sensores e a implementação de uma plataforma para extração e visualização dos dados.

Palavras-Chave

Dados; Internet das coisas; sensores; plataformas; visualização.

Abstract

The data volume is growing day by day and is generated from different sources such as social networks, media, data warehousing, sensors and *IoT* devices from both industry and handheld devices. Different cloud service platforms have been developing methods

and tools for extracting and transforming this data into something of value. The internet of things is in every sector and market, being present in the automotive sector, health, homes, factories, products, public and private safety, the environment, agriculture and food, and other segments. In this article, an IoT device will be developed for sensor-generated data collection and the implementation of a platform for data extraction and visualization.

Keywords

Data; Internet of things; sensors; platforms; visualization.

INTRODUÇÃO

Já se foi a época onde as empresas aguardavam horas ou dias para ver painéis com base em dados antigos e obsoletos. Estudos revelam que cerca 2,2 milhões de terabytes em novos dados estão sendo criados todos os dias no mundo e estima-se que no ano de 2020 serão gerados aproximadamente 1,7 megabytes por segundo de novas informações para cada habitante no mundo (ADNEWS, 2019).

Um outro fator que vem contribuindo para o aumento exponencial do tráfego de dados no mundo são os dispositivos *IoT*, segundo a Cisco (2019) o tráfego IP global anual atingirá 4,8 ZB em 2022. Para se ter uma ideia em 2017 o tráfego IP global foi de 1,5 ZB por ano ou 122 EB por mês.

O crescimento exponencial da quantidade de dados e o aumento de recursos cada vez mais avançados para capturar estes dados vem fazendo com que empresas enfrentem o desafio de dar algum sentido a esta enorme onda de dados brutos. O Hadoop veio como a principal estrutura para lidar com todos estes dados. Até pouco tempo não se tinha recursos ou estruturas para criar aplicativos para o processamento em fluxo. Estes aplicativos permitiram que as empresas descem respostas rápidas a eventos, exemplo disso são os monitoramentos de mídias sociais para saber a aceitação de um determinado produto, análise de sentimentos diante de uma determinada notícia, entre outros. (JAIN, 2017).

Segundo Shields (2018), o espaço público em nuvens é dominado por 5 grandes empresas e de acordo com as previsões da IDC, empresa de inteligência de mercado, espera-se que os gastos globais em nuvem pública aumentem 23,2% chegando a US \$ 160 bilhões em 2018 e nos próximos cinco anos acredita-se que o espaço público em nuvem terá uma taxa de crescimento de 21,9% e os gastos com serviços públicos em nuvem alcancem US \$ 277 bilhões em 2021.

Com o surgimento de várias ferramentas para *Big Data* é fundamental conhecer a aplicabilidade e suas funções, assim como empresas que disponibilizam algumas destas ferramentas, plataformas como AWS (Amazon Web Services) e Microsoft Azure, líderes de mercado, oferecem serviços em nuvem, com soluções e produtos onde é possível construir aplicações com escalabilidade e flexibilidade.

Esse artigo está estruturado por seções que abordam alguns conceitos de Big Data, ferramentas para visualização dos dados, plataformas cloud, internet das coisas, o desenvolvimento de um dispositivo IoT para coleta dos dados e conclusão.

BIG DATA

A etimologia do “*Big Data*” foi usada pela primeira vez por Mashey (1998), ex-cientista chefe aposentado da SGI, onde produziu um deck de slides da SGI intitulado “*Big Data and the Next Wave of InfraStress*”, nesse deck ele demonstrava as várias tendências tecnológicas que seriam esperadas para os próximos anos.

Essa nova era já vem surgindo a algum tempo e tem raízes profundas, o *Big Data* existe há décadas para alguns veteranos da indústria que ao longo dos anos já lidavam com uma grande quantidade de dados transacionais. Com o crescimento de novas tecnologias como a computação móvel, as redes sociais e a computação em nuvem houve um aumento no volume, na velocidade e a variedade de dados. (DHIRAJ; CHAMBERS; MINELLI, 2013).

Du (2018) resume o *Big Data* não se tratando apenas de muitos dados e sim uma prática de descobrir novas informações a partir dos dados existentes e orientar a análise de novos dados. Uma empresa orientada a *Big Data* será mais ágil e competitiva para superar desafios e vencer competições.

Hoje, a tecnologia nos permite coletar dados a uma velocidade impressionante, tanto em termos de volume quanto de variedade. Existem várias fontes que geram dados, mas no contexto de *Big Data*, as fontes principais são as redes sociais, mídia, data warehouse e sensores.

As redes sociais são indiscutivelmente, a principal fonte de todo o *Big Data* que conhecemos hoje e se que proliferaram nos últimos anos. Trata-se, em geral, de dados não estruturados, representados por milhões de publicações nas mídias sociais e outros dados gerados segundo a segundo por meio de interações do usuário na Web em todo o mundo. O aumento do acesso à Internet em todo o mundo tem sido um ato autorrealizável para o crescimento de dados nas redes sociais. (DASGUPTA, 2018)

A mídia é em grande parte o resultado do crescimento das redes sociais e representa os milhões, se não bilhões, de uploads de áudio e vídeos que ocorrem diariamente. Vídeos enviados no YouTube, gravações de música no SoundCloud e imagens postadas no Instagram são exemplos excelentes de mídia, cujo volume continua a crescer de maneira irrestrita. (DASGUPTA, 2018)

As empresas investem há muito tempo em instalações especializadas de armazenamento de dados, comumente conhecidas como *data warehouses*. Um DW é essencialmente uma coleção de dados históricos que as empresas desejam manter e catalogar para fácil recuperação, seja para uso interno ou para fins regulatórios. À medida que os setores passam gradualmente para a prática de armazenar dados em plataformas como Hadoop e NoSQL, mais e mais empresas estão transferindo dados de seus *data warehouses* pré-existentes para algumas das tecnologias mais recentes. E-mails da empresa, registros contábeis, bancos de dados e documentos internos são alguns exemplos de dados DW que agora estão sendo transferidos para plataformas Hadoop ou do tipo Hadoop que utilizam vários *nodes* para fornecer uma plataforma altamente disponível e tolerante a falhas. (DASGUPTA, 2018)

Um fenômeno mais recente no espaço do *Big Data* foi a coleta de dados de dispositivos *IoT* e sensores. Embora sempre existam sensores e indústrias como petróleo e gás usem

sensores de perfuração para medições em plataformas de petróleo há muitas décadas, o advento de dispositivos portáteis, também conhecidos como Internet das Coisas, como Fitbit e Apple Watch, significa que agora cada indivíduo poderia transmitir dados na mesma taxa em que algumas plataformas de petróleo costumavam fazer há alguns anos atrás. (DASGUPTA, 2018)

Tendo como principais características geralmente *Big Data* pode ser explicado de acordo com os três Vs, sendo volume, variedade e velocidade.

- Volume: os dados podem vir de diversas fontes diferentes, como registros de vendas de negócios, os resultados coletados de experimentos científicos ou sensores em tempo real usados na internet das coisas (*IoT*). Os dados podem estar brutos ou pré-processados usando ferramentas de software separadas antes de serem aplicadas as análises. (ROUSE, 2017).
- Variedade: os dados também podem existir em uma grande variedade de tipos de arquivos, sendo dados estruturados, quase estruturados, semiestruturados ou não estruturados. (DHIRAJ; CHAMBERS; MINELLI, 2013).
- Velocidade: refere-se à velocidade na qual os fluxos de dados devem ser analisados. Todo o grande projeto de análise de dados irá ingerir, correlacionar e analisar as fontes de dados e, em seguida, renderizar uma resposta ou resultado com base em uma consulta abrangente. A velocidade também é significativa, uma vez que a análise de dados se expande em campos como aprendizagem de máquinas e inteligência artificial, onde os processos analíticos imitam a percepção ao encontrar e usar padrões nos dados coletados. (ROUSE, 2017).

No entanto existem outras características adicionais como:

- Veracidade: é de grande importância que os dados coletados sejam condizentes com a realidade daquele momento, quando se trata de análise real-time, dados passados não podem ser considerados verídicos no momento que estão sendo analisados. A precisão dentro de um contexto irá impactar os resultados. (HURWITZ et al., 2016). Segundo Rijmenam (2019), a geração dados falhos são inúteis quando coletados, os dados podem causar problemas significativos para as organizações quando incorretos. Garantir que os dados e análises estejam corretos é de extrema importância, principalmente quando a tomada de decisão é de forma automatizada onde nenhum ser humano está mais envolvido. Sistemas de irrigação, por exemplo, falhas nos sensores de humidade de solo podem gerar dados falhos fazendo com que o sistema irrigue de forma incorreta o solo comprometendo assim o plantio e gerando prejuízos aos agricultores.
- Valor: no início do processo de análise saber as perguntas certas é importantíssimo, resultados só terão valor se tiverem sido tratados, analisados e usados para chegar a uma decisão. (TAURION, 2013). Os dados em si não são valiosos. O valor está quando os dados analisados são transformados em informações, nos permitindo adquirir conhecimento e sabedoria. (RIJMENAM, 2014).
- Variabilidade: *Big Data* é extremamente variável, isso significa que uma mesma palavra tem diferentes significados dependendo do contexto e,

portanto, é muito relevante na realização de análises de sentimento. Variabilidade significa que a definição está mudando e rapidamente. Um exemplo são tweets semelhantes, uma palavra pode ter um significado totalmente diferente. Para realizar uma análise de sentimento adequada, os algoritmos precisam decifrar o significado exato de uma palavra em seu contexto. Isso ainda é muito difícil. (RIJMENAM, 2014)

- **Visualização:** Essa é a parte difícil do *Big Data*. Significa tornar essa vasta quantidade de dados compreensível de uma maneira fácil de ler e entender. Com as visualizações corretas, os dados brutos podem ser colocados em uso. As visualizações, claro, não são gráficos comuns ou gráficos de pizza. São gráficos complexos que podem incluir muitas variáveis de dados, permanecendo legíveis e compreensíveis. A visualização pode não ser a tarefa tecnologicamente mais difícil, mas certamente é a mais desafiadora. O uso de um gráfico para contar uma história complexa é muito difícil, mas também extremamente crucial. Felizmente, há mais e mais startups de *Big Data* que se concentram nesse aspecto do desafio. No final, as visualizações farão a diferença. (RIJMENAM, 2014)
- **Volatilidade:** refere-se a quanto tempo os dados são válidos e armazenados. Isso é particularmente importante para análises em tempo real. Ele exige que uma janela de tempo alvo seja determinada para que os analistas possam se concentrar em questões específicas e obter um bom desempenho com a análise. (DU, 2018)

FERRAMENTAS DE VISUALIZAÇÃO

O Tableau é uma ferramenta de BI que possibilita a análise dos dados visualmente. Permite que os usuários criem e distribuam painéis interativos e compartilháveis, descrevendo na forma de gráficos e tabelas as tendências, variações e densidade dos dados. Usado por empresas de diversos setores, pesquisadores acadêmicos e organizações governamentais para análise visual de dados, o Tableau pode conectar-se a arquivos, fontes relacionais e de *Big Data*. Possuindo diversos recursos e um aplicativo de descoberta e exploração de dados permite que o tableau traga resultados em segundos. Alguns recursos exclusivos lhe permitem lidar com diversos cenários, fornecendo soluções para todos os tipos de setores, departamentos e ambientes de dados. Como o Tableau não requer um alto nível de conhecimento em programação, qualquer pessoa que tenha acesso aos dados pode começar a usá-lo para obter valor dos dados, o que é uma vantagem quando se tratando em velocidade de análise. Um outro ponto positivo é de não precisar de uma configuração complexa de software, sua versão para desktop usada pela maioria dos usuários é facilmente instalada e contém todos os recursos necessários para iniciar e concluir análise de dados, tornando-o autossuficiente. (TUTORIALSPPOINT, 2016).

O Power BI é uma ferramenta de visualização de dados e BI que converte dados de diferentes fontes em Dashboards e relatórios de BI. O Power BI suíte fornece vários softwares, conectores e serviços, o Power BI Desktop, o Power BI servisse baseado em SaaS e o Power BI mobile estão disponíveis para diversas plataformas. Esse conjunto de serviços podem ser usados pelos usuários para consumir dados e criar relatórios. O Power BI inclui componentes como Power BI Desktop, usado para criar relatórios e visualizações de dados, enquanto os Serviços de BI (Software como Serviço - SaaS) são usados para publicar os relatórios, o Gateway do Power BI pode ser usado para manter

seus dados atualizados e conectados às fontes de dados locais sem que haja a necessidade de mover os dados, o aplicativo mobile é possível conectar e visualizar os dados de qualquer lugar. (TUTORIALSPPOINT, 2017)

O Power BI suporta uma variedade de fontes de dados muito grande, é possível se conectar a diferentes arquivos simples, banco de dados SQL e nuvem do Azure ou plataformas da web, como Facebook, Salesforce e Google Analytics. É possível também incluir conexão ODBC para conectar a outras fontes de dados. (TUTORIALSPPOINT, 2017)

O WEKA é um software *open source* que fornece ferramentas para pré-processamento de dados, implementação de vários algoritmos de Machine Learning e visualização para que você possa desenvolver técnicas de aprendizado de máquina e aplicá-las a problemas reais de mineração de dados. O WEKA oferece muitos estágios ao lidar com o Big Data, começando com os dados brutos coletados em campo que podem conter valores nulos e campos irrelevantes. Com a ferramenta de pré-processamento é possível limpar estes dados e salvá-los localmente para aplicar os algoritmos. (TUTORIALSPPOINT, 2019)

Dependendo do tipo de modelo que será desenvolvido o WEKA oferece opções de classificação, cluster e associação, cada categoria do WEKA fornece a implementação de vários algoritmos onde você seleciona um algoritmo, define os parâmetros e o executa o conjunto de dados. Vários modelos podem ser aplicados no mesmo conjunto de dados, podendo comparar as saídas de diferentes modelos e selecionar o melhor que atenda aos seus objetivos. (TUTORIALSPPOINT, 2019)

COMPUTAÇÃO EM NUVEM

A computação em nuvem já faz parte da maioria das empresas, independente do seu porte, pois incluem as tecnologias como soluções estratégicas em seus negócios. Vários recursos estão disponibilizados na internet, por exemplo, armazenamento em nuvem. Usuários podem armazenar diversos arquivos como em seus discos locais, porém este armazenamento é fornecido como um serviço de algum provedor de nuvem. Um provedor de nuvem é uma empresa ou grupo de empresas que fornecem serviços em nuvem para pessoas físicas ou jurídicas. Estas empresas hospedam e gerenciam os serviços contratados pelos usuários e são responsáveis por manter este serviço. (MODI, 2019)

Modi (2019) e Wali (2018), destacam algumas vantagens na computação em nuvem, como pagamento conforme o uso, acesso ou presença global o que permite ser acessado de qualquer lugar via internet, recursos ilimitados ou escalabilidade, serviços gerenciados e hardware. Wali (2018) ainda cita outras vantagens sendo economia de custos, economia de espaço, não tem perda de tempo com inatividades causadas por falhas de hardware serão reduzidas.

Existem três principais áreas da computação em nuvem, sendo software como Serviço (SaaS), plataforma como Serviço (PaaS) e infraestrutura como Serviço (IaaS). (SHAH; SARKAR, 2018)

A figura 1 ilustra os níveis de controle que o usuário tem sobre sistemas operacionais, armazenamento, aplicativos, entre outros serviços. Se observamos da esquerda para direita o *User Application* é totalmente controlado pelo usuário e este controle vai

diminuindo gradativamente até que na área SaaS o usuário não tem nenhum tipo de gerenciamento.

O mercado de infraestrutura como serviço (IaaS) vem amadurecendo juntamente com a receita que cresce sem parar. O Gartner, Inc (2019) fez uma projeção em que a receita no mercado de infraestrutura na nuvem deverá aumentar para US \$ 81,5 bilhões até 2022. Para suportar uma ampla variedade de cargas de trabalho o mercado enxerga a AWS e a Microsoft como os dois provedores mais estruturados, o Google vem fazendo progresso constantemente em termos de adoção corporativa, mas está distante em termos de receita anual e de interesse entre os clientes corporativos do Gartner, permanecendo em terceiro lugar.

Figura 1. Nível de Gerenciamento do Usuário

User Application	IaaS Model	PaaS Model	SaaS Model
Application	Application	Application	Application
Data	Data	Data	Data
Runtime (Libraries)	Runtime (Libraries)	Runtime (Libraries)	Runtime (Libraries)
Operating System	Operating System	Operating System	Operating System
Virtualization	Virtualization	Virtualization	Virtualization
Server	Server	Server	Server
Storage	Storage	Storage	Storage
Networking	Networking	Networking	Networking

(SHAH; SARKAR, 2018)

O Amazon Web Services, popularmente conhecido como AWS surgiu em 2006 oferecendo como seu primeiro serviço o *Simple Storage Service* (Serviço de armazenamento simples), abreviado para S3. A ideia era oferecer um conceito de armazenamento de objetos pela web onde qualquer pessoa poderia colocar o objeto, essencialmente qualquer grupo de bytes, desde fotos digitalizadas, backups de arquivos, vídeos, áudios, ou seja, todo e qualquer arquivo que pudesse ser armazenado em um HD local. No começo o S3 era limitado e embora os objetos fossem lidos ou escritos de qualquer lugar, eles eram armazenados apenas na região dos Estados Unidos e não podiam ter mais que 5 gigabytes. (GOLDEN, 2013)

Uma grande parte dos serviços que a AWS oferece se encaixa na definição de infraestrutura como serviço (IaaS), este modelo mais maduro é certamente o mais ofertado, servidores virtualizados e matrizes de armazenamento virtualizado são hospedados em uma rede definida por softwares e com uma infraestrutura completamente isolada para cada cliente como um recurso privado. A criação de recursos na AWS na maioria das vezes começa com a criação do que é chamado de virtual private cloud (VPC). Por sua rede privada ser isolada pode-se hospedar servidores virtuais, HDs virtuais e serviços e produtos completamente gerenciados. A flexibilidade da AWS permite criar um monte de arquiteturas usando vários serviços e utilitários contidos na caixa de ferramentas IaaS. (WILKINS, 2019)

Segundo Wilkins (2019), os serviços de nuvem IaaS são fornecidos com serviços gerenciados que se baseia em serviços de computação, softwares personalizados, armazenamento e rede, fornecendo opções para que a Amazon gerencie e mantenha, assim não precisará ter todo o trabalho.

Em 2008 a Microsoft anuncia o Azure com o nome de *Red Dog* e lançado oficialmente em 2010 com o nome de *Windows Azure* oferecendo funções da Web e banco de dados SQL. Em 2014 a Microsoft alterou o nome de sua plataforma em nuvem para Microsoft Azure. (WALI, 2018)

O Azure oferece suporte a sistemas operacionais como Linux e Windows, linguagens como C#, Python, Java, Node.js e outras, várias ferramentas, plataformas, estruturas e utilitários. A ideia é que o usuário tenha total liberdade em suas escolhas, seja na linguagem, plataforma, sistema operacional, armazenamento, ferramentas e utilitários. O Azure oferece uma série de tecnologias a nível mundial. O Azure possui data centers espalhados pelo mundo todo e estão combinados por regiões, cada região possui vários data centers para garantir uma recuperação mais rápida e eficiente em caso de desastre. (MODI, 2019)

O Azure fornece infraestrutura e serviços para ingerir uma imensa quantidade de transações com processamento, também fornece uma enorme capacidade de armazenamento de dados e uma série de serviços interconectados que podem transmitir dados entre si. Todos esses recursos fazem com que os dados possam ser processados gerando conhecimento e insights significativos. (MODI, 2019)

Em 2008 uma equipe de desenvolvedores do Google uma oferta de plataforma como serviço, o Google App Engine. Nos anos seguintes, o Google lançou vários serviços e recursos como o Google Cloud Storage em 2010, Compute Engine em 2013, Cloud SQL em 2014 e Kubernetes Engine em 2015. Durante esse período, o Google buscou expandir seu domínio em diversas áreas, como análise de dados, gerenciamento de infraestrutura, *IoT* e aprendizado de máquina. Em 2017, o Google havia estabelecido data centers em 39 zonas em 13 regiões. (HUNTER; PORTER; PS, 2019)

Google Cloud Platform possui produtos e serviços que atendem a várias demandas e serviços. Os principais serviços, como o *Compute Engine* e o *Cloud Storage*, permitem praticamente qualquer solução, enquanto muitos serviços especializados, como a *Cloud Vision API*, reduzem possibilitam as equipes enfrentarem espaços de problemas mais específicos. À medida que o Google avança no espaço da nuvem pública, a quantidade de produtos e serviços essenciais e especializados continuam a crescer rapidamente. (HUNTER; PORTER; PS, 2019)

Além das empresas já citadas, existem muitos outros fornecedores de plataforma em nuvem que oferecem serviços voltados para dispositivos *IoT*, quando fomos escolher devemos levar em conta o tipo de projeto, sua aplicabilidade e a complexidade no processo de implantação.

IBM BlueMix – O Bluemix é uma implementação focada em *IoT* da Open Cloud Architecture da IBM, com o Bluemix é possível criar, implementar e gerenciar aplicativos em nuvem voltados para internet das coisas. (SHOVIE, 2016).

IBM Watson – A plataforma IBM Watson é um serviço cloud que pode conectar

dispositivos *IoT* residenciais ou industriais e controlar sensores fornecendo um amplo conjunto de ferramentas integradas e complementares. Com elas é possível processar dados de dispositivos realizando análises em tempo real ou em lote. (IBM CORPORATION, 2019)

ThingSpeak – O ThingSpeak é uma plataforma como serviço (PaaS) de análise de *IoT* que permite agregar, visualizar e analisar fluxos de dados em tempo real. É possível enviar dados dos dispositivos e criar visualizações, bem como enviar alertas. (THINGSPEAK, 2019)

Thingworx – O Thingworx é uma interface do usuário baseada em nuvem para coleta e análise de dados sendo usado, por exemplo, para controle de fábrica. A proprietária da Thingworx em 2015 adquiriu o KepWare, que essencialmente é uma coleção de protocolos e middleware conectados às instalações da produção. Através da combinação dos sistemas os clientes têm acesso aos dados mais recentes da fábrica, assim como tendências e alarmes "outliers". (ELK, 2018)

Dweet – O Dweet.io é basicamente uma API simples que pode ser chamada de qualquer navegador ou aplicativo da Web e pode ser usado sem problemas por qualquer pessoa, graças a sua simplicidade. É realmente fácil armazenar dados de um projeto do Arduino e recuperar esses dados usando outros aplicativos. (SCHWARTZ, 2016)

IoT – INTERNET OF THINGS

O decreto número 9.854, de 25 junho de 2019 art. 2º considera-se:

I – Internet das Coisas – IoT – a infraestrutura que integra a prestação de serviços de valor adicionado com capacidades de conexão física ou virtual de coisas com dispositivos baseados em tecnologias da informação e comunicação existentes e nas suas evoluções, com interoperabilidade; II – Coisas – objetos no mundo físico ou no mundo digital, capazes de serem identificados e integrados pelas redes de comunicação; III – dispositivos – equipamentos ou subconjuntos de equipamentos com capacidade mandatória de comunicação e capacidade opcional de sensoriamento, de atuação, de coleta, de armazenamento e de processamento de dados; e IV – serviço de valor adicionado – atividade que acrescenta a um serviço de telecomunicações que lhe dá suporte e com o qual não se confunde novas utilidades relacionadas ao acesso, ao armazenamento, à apresentação, à movimentação ou à recuperação de informações, nos termos do disposto no art. 61 da Lei nº 9.472, de 16 de julho de 1997. (DIÁRIO OFICIAL DA UNIÃO, 2019, p. 10).

As vantagens da *IoT* abrangem todas as áreas desde estilo de vida à negócios. Aqui está uma lista de algumas das vantagens que a *IoT* tem a oferecer:

- Melhoria do envolvimento do cliente - A análise atual sofre de pontos cegos e falhas significativas na precisão; e como observado, o engajamento permanece passivo. A *IoT* transforma completamente isso para conseguir um engajamento mais rico e eficaz com o público. (TUTORIALSPPOINT,

2016)

- Otimização de Tecnologia - As mesmas tecnologias e dados que melhoram a experiência do cliente também melhoram o uso do dispositivo e auxiliam em melhorias mais potentes para a tecnologia. A *IoT* desbloqueia um mundo de dados funcionais e de campo críticos. (TUTORIALSPPOINT, 2016)
- Resíduos Reduzidos - *IoT* tornam claras as áreas de melhoria. A análise atual nos fornece uma percepção superficial, mas a *IoT* fornece informações reais que levam a um gerenciamento mais eficaz dos recursos. (TUTORIALSPPOINT, 2016)
- Coleta de dados aprimorada - A coleta de dados moderna sofre com suas limitações e seu design para uso passivo. *IoT* separa esses espaços e coloca exatamente onde os humanos realmente querem ir para analisar nosso mundo. Permite uma imagem precisa de tudo. (TUTORIALSPPOINT, 2016)
- Desvantagens, embora a *IoT* ofereça um impressionante conjunto de benefícios, ela também apresenta um conjunto significativo de desafios.
- Segurança - *IoT* cria um ecossistema de dispositivos conectados constantemente, que se comunicam através de redes. O sistema oferece pouco controle apesar de quaisquer medidas de segurança. Isso deixa os usuários expostos a vários tipos de invasores. (TUTORIALSPPOINT, 2016)
- Privacidade - A sofisticação da *IoT* fornece dados pessoais substanciais em detalhes extremos, sem a participação ativa do usuário. (TUTORIALSPPOINT, 2016)
- Complexidade - Alguns acham os sistemas *IoT* complicados em termos de design, implantação e manutenção, devido ao uso de várias tecnologias e um grande conjunto de novas tecnologias de ativação. (TUTORIALSPPOINT, 2016)
- Flexibilidade - Muitos estão preocupados com a flexibilidade de um sistema *IoT* para se integrar facilmente com outro. Eles se preocupam em encontrar-se com vários sistemas conflitantes ou bloqueados. (TUTORIALSPPOINT, 2016)
- Conformidade - A *IoT*, como qualquer outra tecnologia no campo dos negócios, deve obedecer aos regulamentos. Sua complexidade faz com que a questão da conformidade pareça incrivelmente desafiadora quando muitos consideram a conformidade com o software padrão uma batalha. (TUTORIALSPPOINT, 2016)

Existem alguns hardwares de baixo custo disponíveis no mercado para quem queira realizar algum projeto de *IoT*. Dispositivos como Arduino ou Raspberry PI 3 possibilitam a realização de vários projetos. No site dos fabricantes existem vários projetos disponíveis para experimento.

A *IoT* tem aplicativos em todos os setores e mercados. Ele abrange grupos de usuários daqueles que desejam reduzir o uso de energia em suas casas para grandes organizações que desejam otimizar suas operações. Isso não é apenas útil, mas quase crítico em muitos setores, à medida que a tecnologia avança e nos movemos em direção à automação avançada imaginada em um futuro distante. (TUTORIALSPPOINTS, 2016)

DISPOSITIVO IOT

Um dispositivo IoT que está sendo muito utilizado em estudos é um sistema de horta, uma horta para muitas pessoas é algo muito agradável, imagine poder estar em sua casa ou apartamento e enquanto cozinha colher hortaliças fresquinhas, temperos como manjeriço, alecrim, salsinha, cebolinha, entre outros, tudo ali a sua mão fresquinho, porém muitas vezes não temos tempo para regar ou não sabemos quando regar. Um sistema que possa monitorar uma pequena horta e ao mesmo tempo coletar dados para futuras análises seria algo bem interessante. Foi pensando nisso que estamos desenvolvendo o “*Less Water*”, um sistema de irrigação inteligente e sustentável, pois utiliza painéis solares como fonte energia renovável. Neste estudo de caso daremos início ao projeto e iremos abordar alguns aspectos como funcionamento, plataforma, sensores utilizados, montagem e extração dos dados para visualização. O *Less Water* é sistema automatizado e pode ser usado para auxiliar no cultivo de hortaliças, permitindo gerenciar e monitorar a temperatura do ar, umidade do solo e do ar. Geralmente, pequenas hortas são irrigadas de forma manual e sem nenhum parâmetro o que muitas vezes acaba causando a perda das hortaliças.

Os principais componentes usados na montagem do dispositivo foram um Adaptador RoboCore ESP8266 Adapter, uma placa Arduino Uno (Rev3), um sensor de umidade e temperatura DHT22, um módulo WiFi ESP8266 atualizado com ESP-01, um Resistor 10kΩ, um sensor de Umidade de Solo e uma fonte de alimentação para protoboard. A seguir uma breve explicação dos componentes utilizados:

- Sensores – Um dos primeiros parâmetros para a horta é a umidade do solo, Lucietti (2014) destaca a água como o principal componente e que sem ela não ocorre nenhum processo de transformação nos vegetais que vai desde a germinação até todo seu desenvolvimento, sendo assim devemos medir a umidade do solo para que as plantas cresçam bem. No projeto utilizamos um sensor de umidade capacitivo, este sensor pode ser alimentado com tensões que variam de 3,3V à 5,5V e sua tensão de saída vai de 0 à 3V. Uma das vantagens deste sensor está na sua durabilidade, pois por não haver condutores metálicos que ficam em contato direto com o solo não estão sujeitos a corrosão, possui um custo baixo e pode ser facilmente encontrado.

Outro sensor utilizado foi o DHT22, um sensor de umidade e temperatura do ar com tensão de entrada de 3.3V até 6V, com um range de umidade de 0 a 100% e de temperatura de -40 a 80°C e possui uma ótima precisão, este sensor é muito conhecido e possui muitos recursos disponíveis para desenvolvimento.

A temperatura e umidade têm impacto significativo no crescimento das plantas e manter a temperatura e a umidade dentro de certos valores também mantém as hortaliças saudáveis.

O consumo total de água é influenciado pelo tipo de cultura e estágio de desenvolvimento, pelo clima (temperatura, umidade relativa do ar, radiação solar, ventos), pelo tipo de solo, pela cobertura e coloração do solo. O consumo de água pelas plantas pode ser considerado como evapotranspiração = evaporação + transpiração. (LUCIETTI, 2014, on-line)

- Placa MCU e Microcontrolador – para o desenvolvimento do *Less Water* optamos pela placa Arduino UNO R3 por possuir uma plataforma de desenvolvimento simples e de fácil entendimento, uma outra vantagem está na quantidade de material disponível para pesquisa e aplicação.
- Os *protoboard*, também conhecidos como placas de ensaio sem solda nos permiti fazer conexões e alterações sem que haja a necessidade de fazer soldas permanentes em componentes.

Além dos componentes já citados, serão necessários cabos para ligação dos circuitos e sensores, fonte de alimentação, cabo USB para conectar o microcontrolador, alguns adaptadores, caixa para acomodar os circuitos etc.

Antes de iniciarmos a montagem dos circuitos devemos instalar o IDE do Arduino no computador, existe uma outra opção que é o Arduino Web Editor um IDE que faz parte do Arduino Create, mas infelizmente até a data de desenvolvimento do dispositivo este produto não estava disponível no Brasil. Para fazer o download acesse o site arduino.cc. (ARDUINO, 2019)

Após instalado, a tela de desenvolvimento do Arduino é muito simples e é chamada de *sketch* (esboço), na tela de esboço em ferramentas você pode gerenciar as bibliotecas, ver os dados dos sensores pelo monitor serial e configurar no gerenciador de placas o modelo do Arduino ou placas de terceiros. Mais informações podem ser consultadas na aba de ajuda ou pelo site Arduino.cc

Ligação dos sensores

O esquema de ligação do sensor de solo está sendo da seguinte forma:

- O VCC conecta ao 5Volts do Arduino.
- O GND conecta ao GND do Arduino.
- O Sinal conecta ao pino A0 do Arduino.

Como já citado anteriormente para nosso projeto utilizamos o sensor capacitivo, na figura 2 temos o código-fonte usado para testar o funcionamento do sensor de solo.

Figura 2. Código para teste do sensor de solo

```
/*Codigo-fonte para funcionamento do sensor de umidade de solo*/  
void setup(){  
  // Comunicação serial  
  Serial.begin(9600);  
}  
void loop(){  
  int valSolo; // Declaração da variavel  
  valSolo = analogRead(A0); // Define a I/O  
  // Imprime os valores dos sensores na serial  
  Serial.print("Umidade do Solo: ");  
  Serial.println(valSolo);  
  // define tempo entre cada medição  
  delay(2000);  
}
```

Fonte: (KURNIAWAN, 2017)

O esquema de ligação do sensor DHT22 está sendo da seguinte forma:

- O VCC (Pino 1) conecta ao 5Volts do Arduino.
- O GND (Pino 4) conecta ao GND do Arduino.
- O Sinal (Pino 2) conecta ao D9 do Arduino.

É preciso colocar um resistor de 10K Ω entre o VCC e a linha de dados (Pino2) para manter uma comunicação adequada entre o sensor e o microcontrolador.

Na figura 3 o exemplo do código-fonte para leitura do sensor DHT22, diferente do código anterior, foi necessário incluir a biblioteca de sensores DHT. Essa biblioteca pode ser instalada a partir do software Arduino. Vá para o Sketch | Incluir Biblioteca | gerenciar bibliotecas e você receberá uma caixa de diálogo.

Figura 3. Código-fonte para teste do sensor DHT22

```
#include "DHT.h" //inclui a biblioteca DHT  
#define DHTTYPE DHT22 // define o tipo de sensor  
#define DHTPin 9 // DHT Sensor  
DHT dht(DHTPin, DHTTYPE); // Inicializa DHT sensor  
//declaração das variáveis  
float temperatura;  
float umidade;  
void setup() {  
  Serial.begin(9600);  
  pinMode(DHTPin, INPUT);  
  dht.begin();  
}  
void loop() {  
  // Obtém os valores da temperatura e imprime  
  temperatura = dht.readTemperature();  
  Serial.print("Temp: ");  
  Serial.print(temperatura);  
  Serial.print("°C");  
  Serial.print(" | ");  
  // Obtém os valores da umidade e imprime  
  umidade = dht.readHumidity();  
  Serial.print("Umid: ");  
  Serial.print(umidade);  
  Serial.println("%");  
  //Imprime os dados a cada 1 segundo  
  delay(1000);  
}
```

Fonte: (KURNIAWAN, 2017)

Conectividade

Para o envio dos dados via Wi-Fi utilizamos o módulo transceptor Wi-Fi ESP8266-01. O

módulo possui oito pinos disponíveis para conexão com dispositivos de microcontroladores externos. Esse chip pode ser usado como um conector Wi-Fi da placa Arduino, mas também pode ser usado como um NodeMCU independente, usando o IDE do Arduino é possível carregar diretamente um *sketch* na placa ESP8266.

O esquema de ligação modulo ESP-01 está sendo da seguinte forma:

- VCC está conectado no 3.3V do Arduino.
- GND está conectado no GND do Arduino
- RX está conectado no D7 do Arduino.
- TX está conectado no D6 do Arduino.
- Ch-Pd está conectado no 3.3V do Arduino.
- Reset está conectado ao D8 do Arduino.

Na figura 4 temos o código-fonte para teste de conexão da placa ESP-01, da mesma forma que foi feito no DHT é necessário instalar a biblioteca SoftwareSerial para o ESP8622.

Figura 4. Código-fonte para teste de conexão do ESP-01

```
#include <SoftwareSerial.h>
SoftwareSerial EspSerial(6, 7); // Rx, Tx
#define HARDWARE_RESET 8
#define DEBUG true
|
void setup() {
  EspSerial.begin(9600);
  Serial.begin(9600);
  Serial.println("ESP8266 Setup test - use AT coomands");
  connectWiFi();
}
void connectWiFi(void) {
  sendData("AT+RST\r\n", 2000, 0); // reset
  sendData("AT+CWJAP=\"Nome da rede\", \"senha\"\r\n", 2000, 0); //Connect network
  delay(3000);
  sendData("AT+CWMODE=1\r\n", 1000, 0);
  sendData("AT+CIFSR\r\n", 1000, 0); // Show IP Adress
  serial.println("Dispositivo Conectado");
}

String sendData(String command, const int timeout, boolean debug) {
  String response = "";
  EspSerial.print(command);
  long int time = millis();
  while ( (time + timeout) > millis() ) {
    while (EspSerial.available()) {
      // The esp has data so display its output to the serial window
      char c = EspSerial.read(); // read the next character.
      response += c;
    }
  }
  if (debug) {
    Serial.print(response);
  }
  return response;
}
void loop() {
  while (EspSerial.available()) {
    Serial.write(EspSerial.read());
  }
  while (Serial.available()) {
    EspSerial.write(Serial.read());
  }
}
```

Fonte: (ROVAI, 2017)

Fonte energia renovável

Em resumo, a tecnologia fotovoltaica transforma a luz solar em eletricidade. Um sistema fotovoltaico, ou sistema elétrico solar, é composto de várias células solares fotovoltaicas. Uma célula fotovoltaica individual é geralmente pequena, produzindo tipicamente cerca de 1 ou 2 W de energia. Para aumentar a potência das células fotovoltaicas, elas são conectadas para formar unidades maiores chamados módulos. Os módulos, por sua vez, podem ser conectados para formar unidades ainda maiores, chamadas de matrizes, que podem ser interconectadas para produzir mais energia, e assim por diante. (PLANTE, 2014)

Para este projeto utilizado dois painéis fotovoltaico, sendo cada painel com 50W de potência, corrente máxima de 2,76 amperes e voltagem máxima de aproximadamente 18,1 volts. A ligação dos dois painéis foi feita em paralelo para que some as correntes e assim obter aproximadamente 5,5 amperes.

Foi instalado um regenerador de tensão para diminuir a tensão que sai dos painéis para as baterias e circuito, na saída do regenerador para o circuito foi instalado um regulador de tensão de aproximadamente 12 volts, evitando assim possíveis oscilações ou picos de tensão. O sistema é alimentado por duas baterias de 12 volts por 7 amperes ligadas em paralelo para que também somem as correntes e de mais autonomia para a alimentação.

Ligando os componentes

Após testados todos os componentes vamos unificar tudo, na tabela 1 temos uma sequência de imagens do processo de instalação dos painéis, baterias e dispositivo, abaixo também é possível conferir o esquema de ligação dos sensores no Arduino.

No Arduino:

- O Sinal do sensor de solo conecta ao pino A0 do Arduino
- O Sinal do sensor DHT22 (Pino 2) conecta ao D9 do Arduino.
- No ESP-01 o RX está conectado no D7 do Arduino.
- No ESP-01 o TX está conectado no D6 do Arduino.
- No ESP-01 o Reset está conectado ao D8 do Arduino.

Na fonte de alimentação:

- O VCC do sensor de solo conecta ao 5Volts da protoboard.
- O GND do sensor de solo conecta ao GND da protoboard.
- O VCC (Pino 1) do DHT22 conecta ao 5Volts da protoboard.
- O GND (Pino 4) do DHT22 conecta ao GND da protoboard.

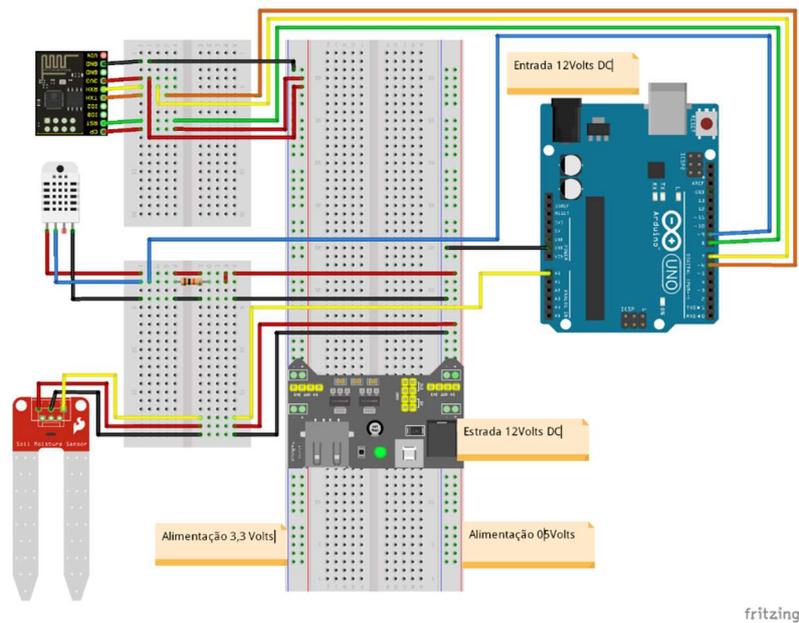
É preciso colocar um resistor de $10K\Omega$ entre o VCC da protoboard e a linha de dados (Pino2) do DHT22 para manter uma comunicação adequada entre o sensor e o microcontrolador.

- VCC do ESP-01 está conectado no 3.3V da protoboard.
- GND do ESP-01 está conectado no GND da protoboard.
- Ch-Pd do ESP-01 está conectado no 3.3V da protoboard.
- O GND do Arduino deve ser ligado do GND do protoboard.

Para o desenho do esquema utilizamos o aplicativo Fritzing, um diagrama de Fritzing é uma representação gráfica da aparência do circuito montado, uma das vantagens é que possui um símbolo semelhante ao componente físico o que facilita sua montagem.

Na figura 5 temos o diagrama completo do nosso projeto IoT.

Figura 5. Esquema Inicial do Projeto



Definindo uma plataforma big data para IoT

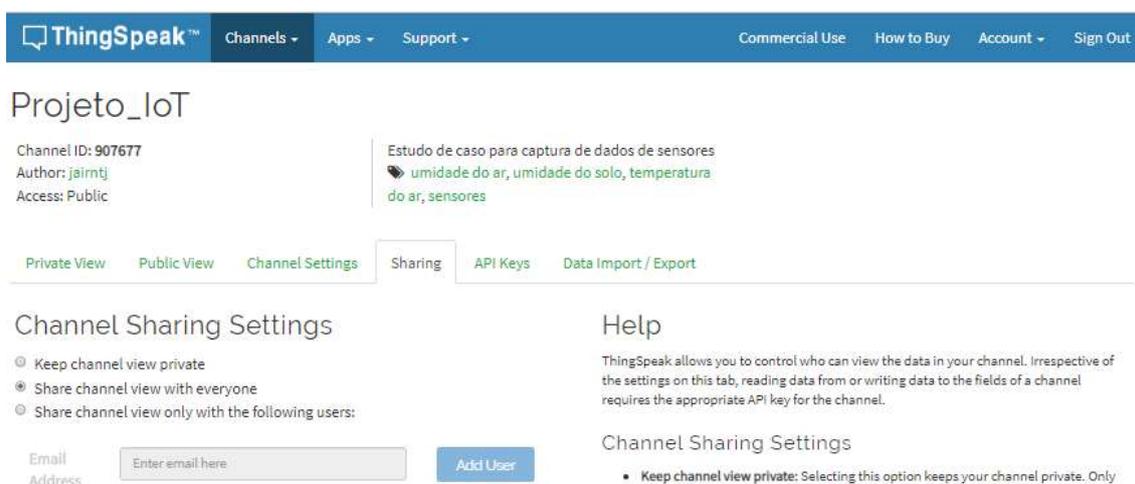
Após o dispositivo montado e testado vamos dar sequência ao nosso projeto implementando uma plataforma Big Data. Foi escolhida uma Plataforma como Serviço (PaaS) para este projeto, pois desta forma podemos nos concentrar muito mais na funcionalidade de seus aplicativos. Entre as empresas citadas anteriormente optamos pela plataforma cloud ThingSpeak para projetos IoT pela sua facilidade de uso com Arduino, por possuir vários recursos on-line, além de não requerer cartão de crédito para registro gratuito e uso acadêmico. Como certeza é a melhor maneira de começar a explorar o mundo da internet das coisas.

As etapas básicas consistem em:

- Abrir o site <https://thingspeak.com> para registro e inscrição.
- Criar um canal selecionando Channels.

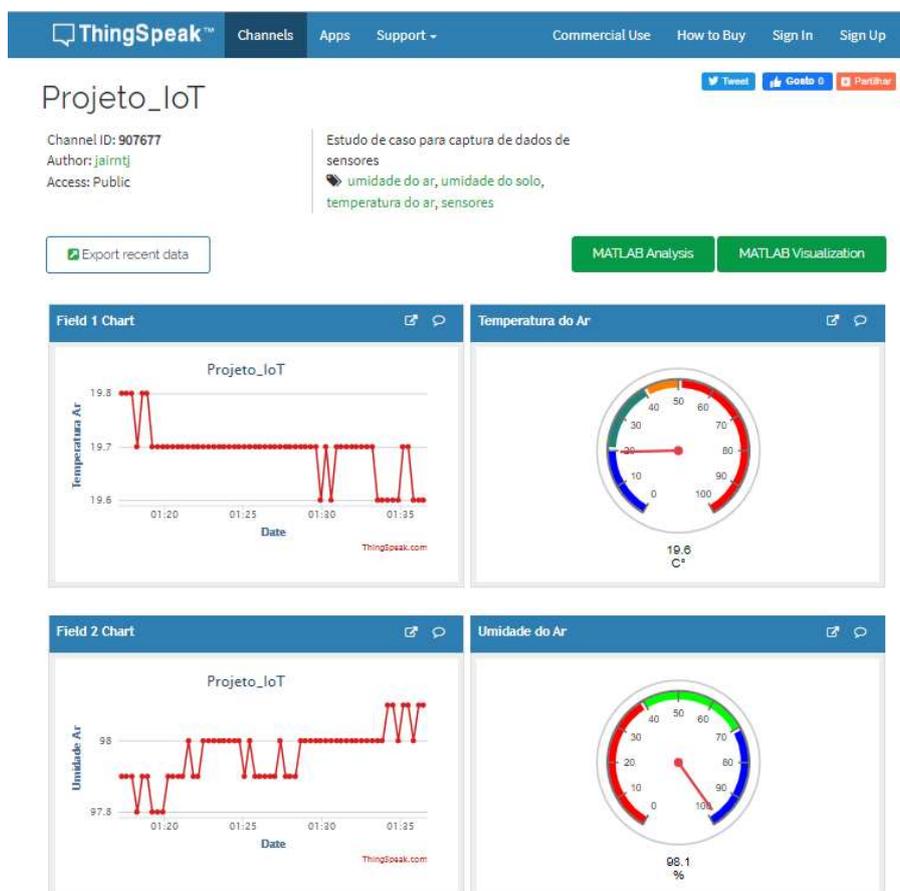
Após concluída as etapas acima devemos anotar o Write API Key, pois esta chave é necessária para gravar os dados no canal e deverá ser inserida em statusChWhiteKey conforme mostrado na figura 6.

Figura 8. Tela de opção para canal privado ou público



Na plataforma em channels é possível visualizar vários canais públicos ou buscar um canal específico através do ID, que neste caso cadastramos como jairmtj e Projeto_IoT. Abaixo tela de visualização do canal público, na opção *Export recent data* o usuário pode exportar os 100 últimos dados armazenados.

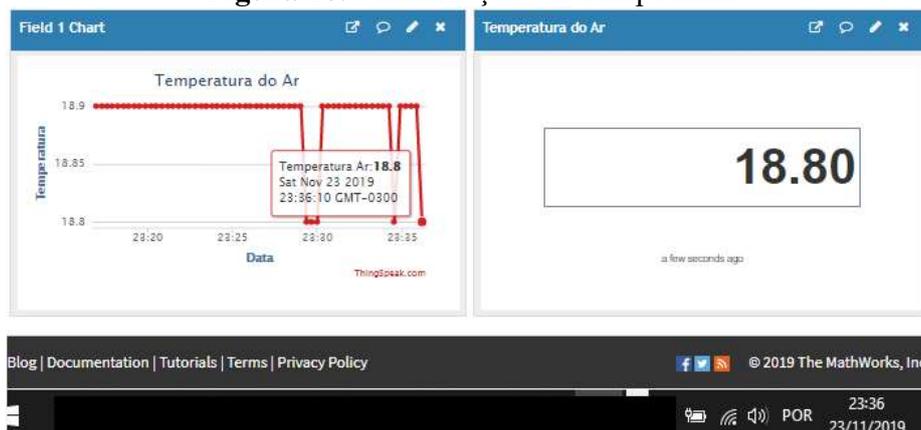
Figura 9. Tela visualização pública



Extraindo os dados para visualização

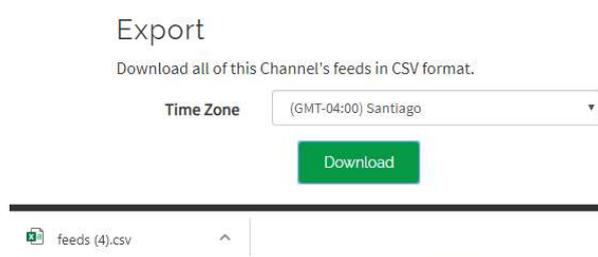
Existe duas maneiras de visualização dos dados, sendo uma em tempo real e outra em lote. Nas visualizações em tempo real os dados são gerados pelos sensores instalados, coletados pelo microcontrolador e transmitidos para a plataforma ThingSpeak pelo módulo Wi-Fi ESP-01. A plataforma por sua vez, transmite os dados através de suas API para visualização e ao mesmo tempo armazena esses dados. Na figura 10 temos um exemplo de dados sendo visualizados em tempo real.

Figura 10. Visualização em Tempo Real



Para extrair os dados em lote, devemos ir na aba *Data Import/Export* e em *Export* devemos escolher o fuso horário, quando houver e clicar em *download*. Devido a não mudança do horário de verão neste ano de 2019, escolhemos o fuso horário de (GMT-04:00) Santiago para que as datas e horários correspondam à realidade. Na figura 11, exemplo da exportação dos dados onde é gerado um arquivo com nome feeds.csv.

Figura 11. Exportação dos Dados



Durante o período de 19/11/2019 às 00:00:10 até o dia 23/11/2019 às 23:59:58, foram geradas 21688 linhas de dados, cada linha possui data/hora, Id e valores dos sensores (figura 12). Dados como *latitude*, *longitude*, *elevation* e *status* não foram validados nesta coleta e deverão ser descartados no processo de transformação.

Figura 12. Arquivo feeds.csv

	A
1	created_at,entry_id,field1,field2,field3,latitude,longitude,elevation,status
2	2019-11-19 00:00:10 -03,1,19.30,92.20,272,,,,
3	2019-11-19 00:00:49 -03,2,19.30,92.30,272,,,,
...	
21686	2019-11-23 23:59:00 -03,21685,18.80,99.90,270,,,,
21687	2019-11-23 23:59:20 -03,21686,18.80,99.90,270,,,,
21688	2019-11-23 23:59:39 -03,21687,18.80,99.90,270,,,,
21689	2019-11-23 23:59:58 -03,21688,18.80,99.90,271,,,,

Carregando os Dados

Para gerar as visualizações dos dados extraídos iremos a utilizar o Power BI e o WEKA para crias as visualizações, primeiramente iremos carregar os dados. No Power BI em obter dados, selecione a opção desejada, no nosso caso selecionaremos texto/csv.

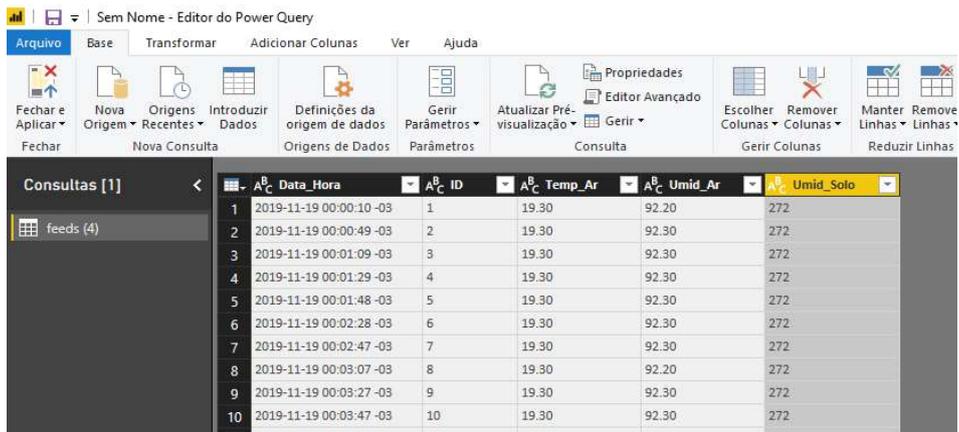
Na figura 13 temos uma pré-visualização dos dados obtidos.

Figura 13. Dados Obtidos para Análise

created_at	entry_id	field1	field2	field3	latitude	longitude	elevation	status
19/11/2019 00:00:10 -03:00	1	1930	9220	272				
19/11/2019 00:00:49 -03:00	2	1930	9230	272				
19/11/2019 00:01:09 -03:00	3	1930	9230	272				
19/11/2019 00:01:29 -03:00	4	1930	9230	272				
19/11/2019 00:05:05 -03:00	14	1930	9230	272				
19/11/2019 00:05:25 -03:00	15	1930	9220	272				
19/11/2019 00:05:45 -03:00	16	1940	9230	272				
19/11/2019 00:06:05 -03:00	17	1940	9230	272				
19/11/2019 00:06:25 -03:00	18	1930	9220	272				
19/11/2019 00:06:44 -03:00	19	1930	9220	272				
19/11/2019 00:07:04 -03:00	20	1940	9220	271				

Antes de carregarmos os dados para gerar as visualizações devemos editar descartando os dados nulos ou em branco. Na tela de pré-visualização clique em editar e faça as modificações necessárias. Na figura 14 temos as colunas latitude, longitude, elevation e status removidas. Também renomeamos as colunas dos dados que vamos utilizar.

Figura 14. Dados Editados

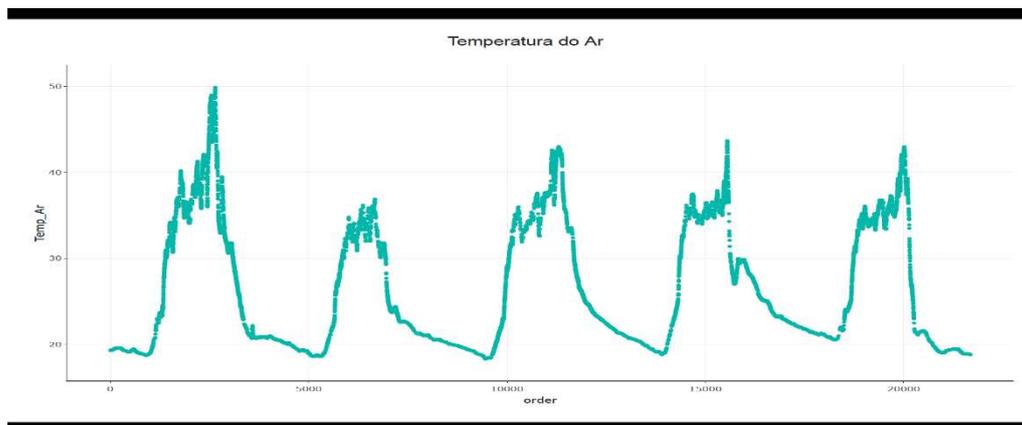


	A ^B _C Data_Hora	A ^B _C ID	A ^B _C Temp_Ar	A ^B _C Umid_Ar	A ^B _C Umid_Solo
1	2019-11-19 00:00:10-03	1	19.30	92.20	272
2	2019-11-19 00:00:49-03	2	19.30	92.30	272
3	2019-11-19 00:01:09-03	3	19.30	92.30	272
4	2019-11-19 00:01:29-03	4	19.30	92.30	272
5	2019-11-19 00:01:48-03	5	19.30	92.30	272
6	2019-11-19 00:02:28-03	6	19.30	92.30	272
7	2019-11-19 00:02:47-03	7	19.30	92.30	272
8	2019-11-19 00:03:07-03	8	19.30	92.20	272
9	2019-11-19 00:03:27-03	9	19.30	92.30	272
10	2019-11-19 00:03:47-03	10	19.30	92.30	272

Gerando gráficos

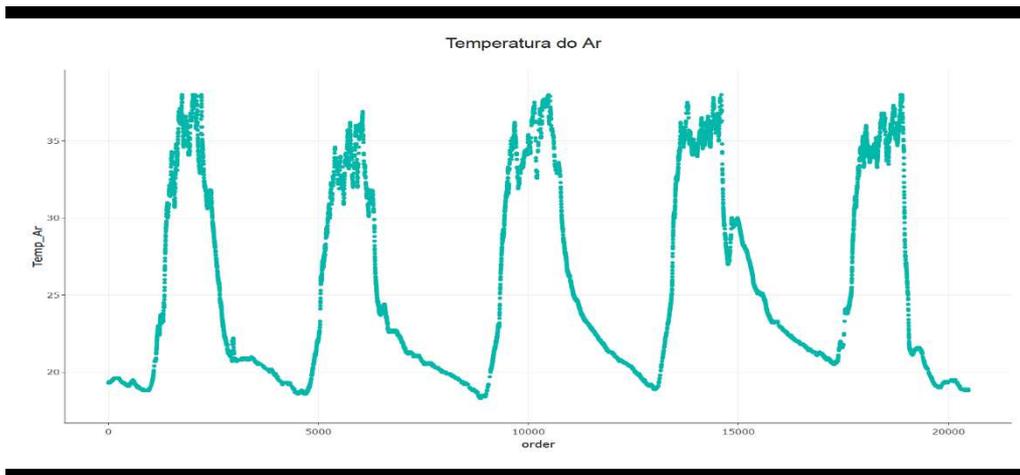
A primeira visualização gerada foi a temperatura do ar e já podemos ver que existe algo de errado com os valores, pois segundo Pegorim (2019) a maior temperatura registrada na cidade de São Paulo deste 1961 até 2018 foi de 37.8°C.

Gráfico 1. Temperatura do Ar



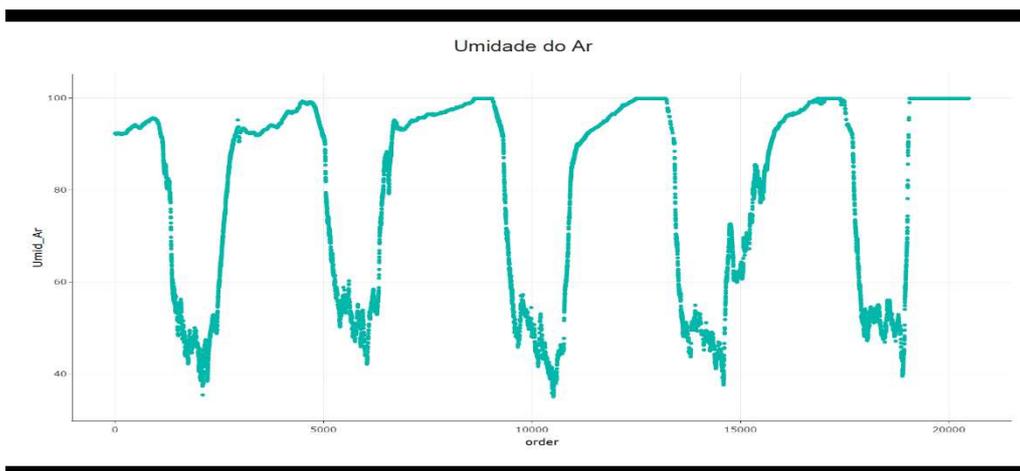
Essas temperaturas foram vistas também em tempo real, percebemos que durante a noite os valores correspondiam ao Climatempo e durante o dia os valores eram bem superiores aos do site, pois o sol estava batendo diretamente na caixa do circuito o que causava o aquecimento dela interferindo assim nos valores gerados pelo sensor que estava dentro da caixa. Para corrigir esta falha retiramos o sensor de temperatura e umidade de dentro da caixa o que causou uma redução dos valores obtidos. Com base nas informações do Climatempo, descartamos os valores acima de 38°C, gerando o gráfico com valores mais próximos da realidade.

Gráfico 2. Valores excluídos



No gráfico 3 podemos visualizar os dados da umidade do ar.

Gráfico 3. Umidade do Ar



Ao juntarmos os dois valores notamos que um é o contrário do outro, ou seja, quanto maior a temperatura do ar, menor será a umidade do ar ou vice-versa.

Gráfico 4. Comparação entre temperatura e umidade

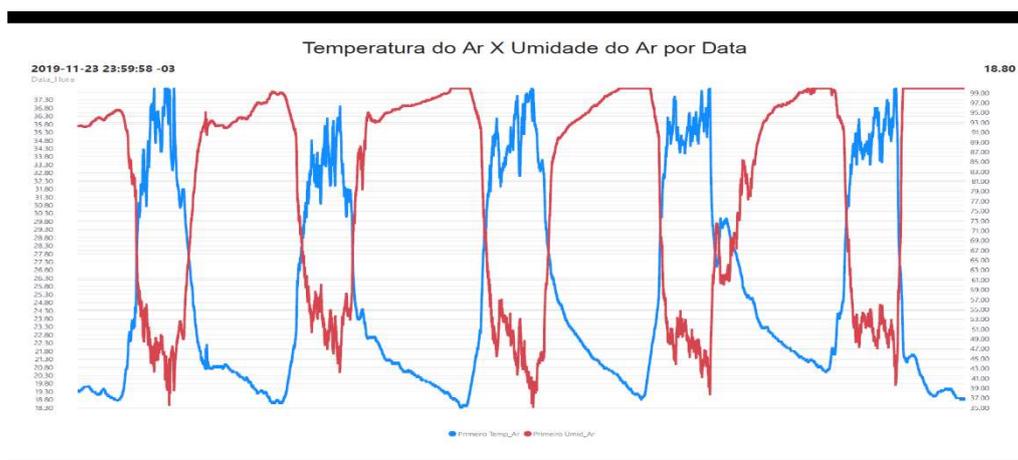
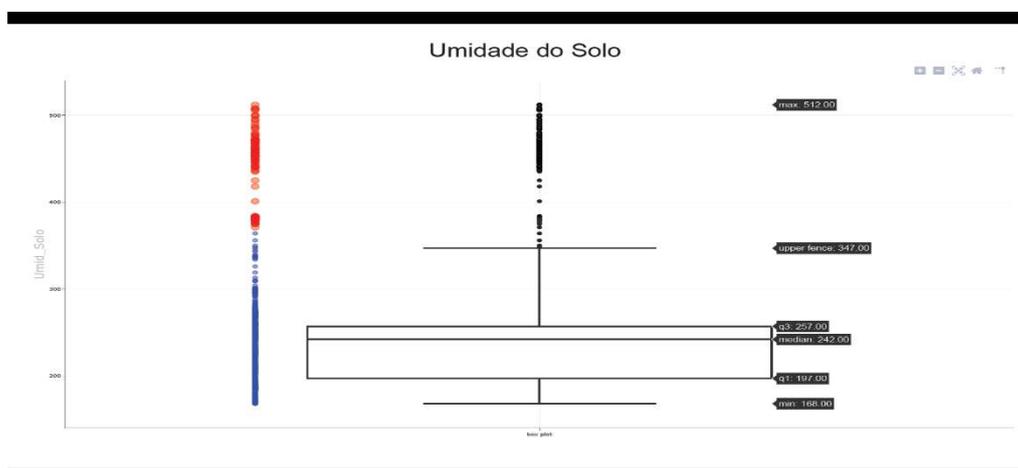


Gráfico 5. Umidade do solo



O gráfico 5 geramos um box plot e foi possível identificar alguns *Outliers* que estão acima do limite superior, os valores obtidos são:

- Limite Inferior – 168.00
- Primeiro Quartil Q_1 – 197.00
- Segundo Quartil ou Mediana Q_2 – 242.00
- Terceiro Quartil Q_3 – 257.00
- Limite Superior – 347.00

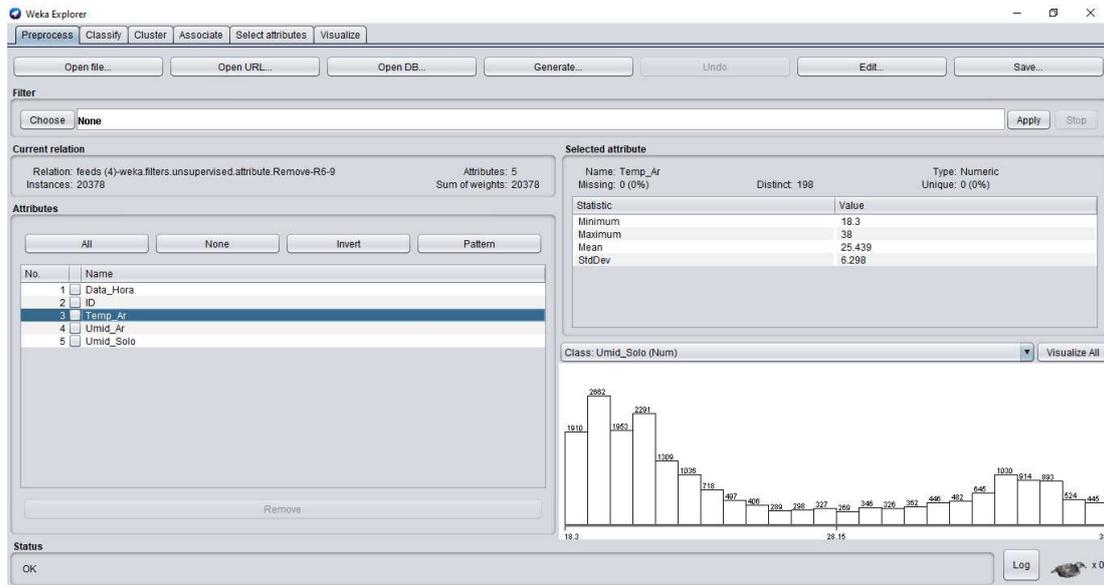
Ainda é cedo para estipular um padrão para analisar a umidade do solo, caso tivéssemos pelo menos 30 dias de dados coletados, poderíamos criar um padrão para solo seco, úmido e molhado. Após removermos os *Outliers* abaixo os novos valores:

- Limite Inferior – 167.00
- Primeiro Quartil Q_1 – 197.00
- Segundo Quartil ou Mediana Q_2 – 234.00
- Terceiro Quartil Q_3 – 250.00

- Limite Superior – 326.00

Para gerar as visualizações no WEKA devemos pré-processar os dados da mesma forma que fizemos no Power BI. Na figura 19 os dados já foram tratados, foram excluídas as colunas em branco, as temperaturas acima de 38°C, os *Outliers* do sensor de umidade de solo e renomeado as colunas.

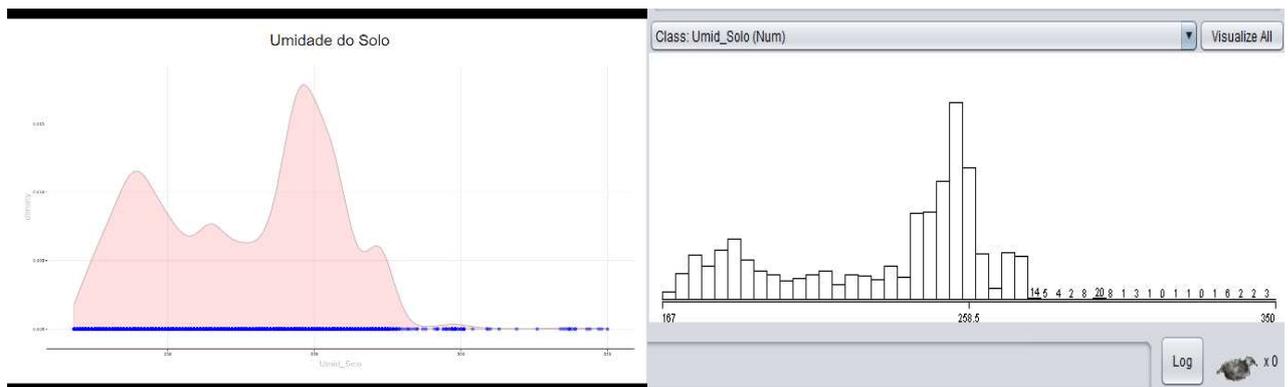
Figura 15. Tela de pré-processamento com dados tratados



Uma das vantagens do WEKA é que na tela de pré-processamento ele já nos traz informações como mínima, máxima, média, desvio padrão e a visualização de um histograma.

Segundo Wilke (2019), os histogramas são uma opção popular de visualização desde pelo menos o século 18, em parte porque eles são facilmente gerados à mão. Com o crescimento das tecnologias e o poder computacional os histogramas vêm sendo substituídos cada vez mais por gráficos de densidade em que tentamos visualizar a distribuição de probabilidade subjacente dos dados desenhando uma curva contínua. Na figura 20 comparamos um gráfico de densidade gerado pelo Power BI a partir dos dados coletados dos sensores e um histograma gerado pelo WEKA com os mesmos dados e podemos notar a semelhança entre eles.

Figura 16. Densidade Power BI e Histograma WEKA



CONCLUSÕES

A presente pesquisa se propôs a desenvolver um dispositivo IoT com o objetivo de coletar dados gerados por sensores, a implementação de arquitetura *big data* ou plataformas cloud como serviços (PaaS) e gerar visualizações em tempo real ou em lote para análises.

No processo de desenvolvimento do *Less Water*, pesquisei sobre alguns microcontroladores e dispositivos para conectividade, pois para ser considerado um dispositivo IoT ele precisa estar conectado. O que pude perceber foi que um simples dispositivo gera uma quantidade de dados sem precedentes, como já dito anteriormente em apenas 4 dias foram geradas 21688 linhas de dados contendo data/hora, ID e dados de 3 sensores, uma média de 5422 dados por dia ou para ser mais exato uma linha de dados a cada 16 segundos. Em um ano este único dispositivo pode gerar até 1.979.000 linhas de dados ou 78,8Mbytes por ano. A questão é, haverá infraestrutura para atender toda esta demanda?

Isso irá depender de como e o que queremos consumir como dados. Várias plataformas como serviços (PaaS) oferecem serviços voltados para o universo IoT, escolhi a Plataforma ThingSpeak por ter sido citada em diversas bibliografias, pela facilidade de implantação e para estudos acadêmicos se torna atrativa por não exigir cartão de crédito. Nesse projeto de pesquisa ela atendeu bem aos meus objetivos, disponibilizando visualizações dos dados em tempo real e armazenando ao mesmo tempo.

Com relação as ferramentas de visualização e análise, foi possível detectar algumas falhas nos sensores de temperatura e de umidade de solo, no caso da temperatura os valores estavam bem acima do normal, causando assim uma ação corretiva, já no sensor de umidade do solo o gráfico do box plot nos permitiu criar uma escala para as condições do solo e isso poderá ajudar em trabalhos futuros quando for implantado um sistema de irrigação em que com base nos dados dos sensores decida quando irrigar.

Em outra oportunidade de pesquisa verificar se é possível incluir algum algoritmo de aprendizado de máquina em que através de dados coletado possa decidir se deve irrigar ou não.

REFERÊNCIAS BIBLIOGRÁFICAS

ADNEWS, Redação. **5 plataformas de gestão de dados que todo profissional de marketing deveria conhecer**. Disponível em: <<https://adnews.com.br/adtech/5-plataformas-de-gestao-de-dados-que-todo-profissional-de-marketing-deveria-conhecer/>>. Acesso em: 04 jul. 2019.

ARDUINO. **ARDUINO UNO REV3**. [S. l.], 6 nov. 2019. Disponível em: <https://store.arduino.cc/usa/arduino-uno-rev3>. Acesso em: 27 out. 2019.

CISCO, Cisco. **Cisco Visual Networking Index: Forecast and Methodology, 2017-2022**. 2019. Disponível em: <<https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-741490.html>>. Acesso em: 04 julho.2019.

DASGUPTA, Nataraj. **Practical Big Data Analytics: Hands-on techniques to implement enterprise analytics and machine learning using Hadoop, Spark, NoSQL and R**. United Kingdom: Packt Publishing Limited, 2018. 412 p. ISBN 9781783554393. Disponível em:

<https://learning.oreilly.com/library/view/practical-big-data/9781783554393/>. Acesso em: 17 ago. 2019.

DHIRAJ, Ambiga; CHAMBERS, Michele; MINELLI, Michael. **Big Data, Big Analytics: Emerging Business Intelligence and Analytic Trends for Today's Businesses**. New Jersey: John Wiley & Sons, Inc., 2013. 224 p. (ISBN: 9781118147603). Disponível em: <<https://learning.oreilly.com/library/view/big-data-big/9781118239155/>>. Acesso em: 19 jul. 2019.

DIÁRIO OFICIAL DA UNIÃO. DECRETO nº N° 9.854, de 25 de junho de 2019. Art. 84, caput, inciso VI, alínea “a”, da Constituição. **Institui o Plano Nacional de Internet das Coisas e dispõe sobre a Câmara de Gestão e Acompanhamento do Desenvolvimento de Sistemas de Comunicação Máquina a Máquina e Internet das Coisas.**, Brasília, p. 10-11, 26 jun. 2019. Disponível em: <http://pesquisa.in.gov.br/imprensa/jsp/visualiza/index.jsp?data=26/06/2019&jornal=515&pagina=10&totalArquivos=149>. Acesso em: 26 nov. 2019.

DU, Dayong. **Apache Hive Essentials**. 2. ed. rev. Birmingham, UK.: Packt Publishing Ltd., 2018. 210 p. ISBN 9781788995092. Disponível em: <https://learning.oreilly.com/library/view/apache-hive-essentials/9781788995092/>. Acesso em: 31 ago. 2019.

ELK, Klaus. **Embedded Software for the IoT**. 3ª. Ed. Berlim, Alemanha: Walter de Gruyter Inc., 2018. ISBN 978-1-5474-0104-8. *E-book*.

GARTNER, INC. Magic Quadrant for Cloud Infrastructure as a Service, Worldwide. **Magic Quadrant for Cloud Infrastructure as a Service, Worldwide**, [s. l.], 16 jul. 2019. Disponível em: <https://www.gartner.com/doc/reprints?id=1-10BCYZ24&ct=190725&st=sb>. Acesso em: 11 out. 2019.

GOLDEN, Bernard. **Amazon Web Services™ For Dummies®**. NJ, USA: John Wiley & Sons, Inc., 2013. 346 p. ISBN 978-1-118-65198-8. *E-book*.

HUNTER, Ted; PORTER, Steven; PS, Legorie Rajan. **Building Google Cloud Platform Solutions**. Birmingham, UK.: Packt Publishing Ltd., 2019. ISBN 978-1-83864-743-8. *E-book*.

HURWITZ, Judith et al. **Big Data para leigos**. 1ª. ed. Rio de Janeiro: Alta Books, 2016. 302 p.

IBM CORPORATION. **IBM Watson IoT Platform**. [S. l.], 2019. Disponível em: https://www.ibm.com/support/knowledgecenter/en/SSQP8H/IoT/kc_welcome.htm. Acesso em: 12 nov. 2019.

JAIN, Ankit. **Mastering Apache Storm**. Birmingham, UK.: Packt Publishing Ltd., 2017. ISBN 978-1-78712-563-6. Disponível em: <https://learning.oreilly.com/library/view/mastering-apache-storm/9781787125636>. Acesso em: 19 set. 2019.

KURNIAWAN, Agus. **Intelligent IoT Projects in 7 Days**. Birmingham, UK.: Packt Publishing Ltd., 2017. ISBN 978-1-78728-642-9. Disponível em: <https://learning.oreilly.com/library/view/intelligent-IoT-projects/9781787286429/>. Acesso em: 25 out. 2019.

LUCIETTI, Donato. **Irrigação das Hortaliças**. Santa Catarina, 2014. Disponível em: <http://cultivehortaorganica.blogspot.com/2014/01/irrigacao-das-hortalicas.html>. Acesso em: 25 out. 2019.

MASHEY, John. **Big Data and the Next Wave of InfraStress**. 1998. Disponível em: <http://static.usenix.org/event/usenix99/invited_talks/mashey.pdf>. Acesso em: 08 jul. 2019

MODI, Ritesh. **Azure for Architects: Implementing cloud design, DevOps, containers, IoT, and serverless solutions on your public cloud**. 2ª. ed. Birmingham, UK.: Packt

Publishing Ltd., 2019. 536 p. ISBN 9781789614503. Disponível em: <https://learning.oreilly.com/library/view/azure-for-architects/9781789614503/>. Acesso em: 14 out. 2019.

PEGORIM, Josélia. **Novo recorde de calor em São Paulo**. [S. l.]: ClimaTempo, 9 jan. 2019. Disponível em: <https://www.climatempo.com.br/noticia/2019/01/09/novo-recorde-de-calor-em-sao-paulo-0898>. Acesso em: 23 nov. 2019.

PLANTE, Russell H. **Solar Energy, Photovoltaics, and Domestic Hot Water**. Califórnia, EUA: Elsevier Editora LTDA, 2014. ISBN 978-0-12-420155-2. Disponível em: <https://learning.oreilly.com/library/view/solar-energy-photovoltaics/9780124201552/>. Acesso em: 19 nov. 2019.

RIJMENAM, Mark Van. **Think Bigger: Developing a Successful Big Data Strategy for Your Business**. New York: Amacom, 2014. 288 p. (ISBN-13: 978-0-8144-3415-4). Disponível em: <https://learning.oreilly.com/library/view/think-bigger/9780814434154/>. Acesso em: 02 jul. 2019.

ROUSE, Margaret. **What is big data?** Disponível em: <http://searchcloudcomputing.techtarget.com/definition/big-data-Big-Data>>. Acesso em: 20 ago. 2019.

ROVAI, Marcelo José. **“IoT feito fácil”**: Conectando coisas desde qualquer canto do mundo! [S. l.], 2017. Disponível em: <https://mjrobot.org/2017/12/21/IoT-feito-facil-conectando-coisas-desde-qualquer-canto-do-mundo/>. Acesso em: 6 nov. 2019.

SARKAR, Aurobindo. **Learning Spark SQL: Architect streaming analytics and machine learning solutions**. Birmingham - Mumbai: Packt Publishing, 2017. 452 p. Disponível em: <https://www.safaribooksonline.com/library/view/learning-spark-sql/9781785888359/7b6bdf28-8f58-47f5-a94c-1fb7d7261973.xhtml>>. Acesso em: 26 fev. 2018.

SCHWARTZ, Marco. **Internet of Things with Arduino Cookbook**. Birmingham, UK.: Packt Publishing Ltd., 2016. ISBN 978-1-78528-658-2. Disponível em: <https://learning.oreilly.com/library/view/internet-of-things/9781785286582/>. Acesso em: 13 nov. 2019.

SHAH, Amit; SARKAR, Aurobindo. **Learning AWS: Design, build, and deploy responsive applications using AWS Cloud components**. 2ª. ed. Birmingham, UK.: Packt Publishing Ltd., 2018. ISBN 978-1-78728-106-6. Disponível em: <https://learning.oreilly.com/library/view/learning-aws-/9781787281066>. Acesso em: 15 out. 2019.

SHIELDS, Anne. **Is Azure Responsible for Microsoft's All-Time Highs?: Amazon and Microsoft are leading the public cloud space**. 2018. Disponível em: <https://marketrealist.com/2018/03/is-azure-responsible-for-microsofts-all-time-highs>>. Acesso em: 09 mar. 2019.

SHOVIE, John C. **Raspberry Pi IoT Projects: Prototyping Experiments for Makers**. Washington, USA: Apress, 2016. ISBN 978-1-4842-1377-3. *E-book*.

TAURION, Cezar. **Big Data**. Rio de Janeiro: Brasport, 2013. 102 p.

THINGSPEAK. **ThingSpeak for IoT Projects**. [S. l.], 2019. Disponível em: <https://thingspeak.com>. Acesso em: 12 nov. 2019.

TUTORIALSPPOINT (org.). **Internet of Things (IoT) Tutorial**. 2016. Disponível em: https://www.tutorialspoint.com/internet_of_things/index.htm>. Acesso em: 07 maio 2018.

TUTORIALSPPOINT (org.). **Power BI Tutorial**. [S. l.], 2017. Disponível em: https://www.tutorialspoint.com/power_bi/index.htm. Acesso em: 6 out. 2019.

TUTORIALSPPOINT (org.). **Tableau Tutorial**. [S. l.], 2016. Disponível em: <https://www.tutorialspoint.com/tableau/index.htm>. Acesso em: 6 out. 2019.

TUTORIALSPPOINT (org.). **WEKA Tutorial**. [S. l.], 2019. Disponível em: <https://www.tutorialspoint.com/weka/index.htm>. Acesso em: 24 nov. 2019.

WALI, Mohamed. **Learn Microsoft Azure: Build, manage, and scale cloud applications using the Azure ecosystem**. Birmingham, UK.: Packt Publishing Ltd., 2018. 354 p. ISBN 9781789617580. Disponível em: <https://learning.oreilly.com/library/view/learn-microsoft-azure/9781789617580/>. Acesso em: 14 out. 2019.

WILKE, Claus O. **Fundamentals of Data Visualization: A Primer on Making Informative and Compelling Figures**. Califórnia, EUA: O'Reilly Media, Inc., 2019. ISBN 9781492031079. Disponível em: <https://learning.oreilly.com/library/view/fundamentals-of-data/9781492031079>. Acesso em: 22 nov. 2019.

WILKINS, Mark. **Learning Amazon Web Services (AWS): A Hands-On Guide to the Fundamentals of AWS Cloud**. [S. l.]: Addison-Wesley Professional, 2019. ISBN 9780135301104. Disponível em: <https://learning.oreilly.com/library/view/learning-amazon-web/9780135301104/>. Acesso em: 11 out. 2019.