CONSTRUÇÃO DE UM AMIENTE ANALITICO PARA ENSINO E PESQUISA BASEADO EM BIG DATA E INTERNET DAS COISAS

Abram Grossmann (IC) e Leandro Augusto da Silva (Orientador)

Apoio: PIBIT Mackpesquisa

RESUMO

Por ser uma área interdisciplinar, a Internet das Coisas apresenta grandes desafios para o aprendizado. No entanto, ela já faz e continuará fazendo parte do dia a dia e, por isso, exige profissionais qualificados para o avanço dos projetos nessa área. Além de adquirir conceitos teóricos, os alunos precisam colocar o conhecimento em prática. Esta aprendizagem prática visa proporcionar um meio de fácil assimilação ao aluno e que possa espelhar situações reais de implementação. Este trabalho apresenta uma metodologia de aprendizagem da Internet das Coisas baseada no desenvolvimento de ambientes que possibilitem ao aluno colocar o conhecimento teórico em prática em um cenário de fácil assimilação. Espera-se que o aluno seja capaz de compreender o processo de desenvolvimento de projetos de Internet das Coisas e as tecnologias envolvidas. A metodologia proposta é composta por 5 etapas. O aluno analisa o ambiente de desenvolvimento, define o tipo de implementação a ser realizada, desenvolve o hardware, o software e os documentos do projeto. A arquitetura de dados em conjunto com a metodologia permitem ao aluno utilizar e propor diversos tipos de ambientes de desenvolvimento, controladores e aplicações web, sendo muito flexíveis para o aprendizado. A implantação do controle de temperatura foi realizada em ambiente aquático. A metodologia proposta mostrou-se eficiente para o desenvolvimento deste projeto, podendo ser aplicada no ensino de Internet das Coisas em instituições de ensino.

PALAVRAS-CHAVE: Internet das Coisas, Big Data e Monitoramento.

ABSTRACT

Being an interdisciplinary area, Internet of Things presents great challenges to learning. However, it already is and will continue to be part of the daily life and thus requires qualified professionals to advance projects in this area. Apart from acquiring theoretical concepts, students need to put knowledge into practice. This practical learning aims to provide a means of easy assimilation to the student and that can mirror real situations of implementation. This work presents an Internet of Things learning methodology based on the development of environments that enable the student to put theoretical knowledge into practice in a scenario of easy assimilation. It is expected that the student will be able to understand the process of developing Internet of Things projects and the technologies involved in it. The proposed methodology is composed of 5 steps. The student analyzes the development environment, defines the type of implementation to be carried out, develops the hardware, the software and documents of the project. The data architecture together with the methodology allow the student to use and propose various types of development environments, controllers and web applications, being very flexible for learning. The implementation of temperature control was carried out in an aquarium environment. The proposed methodology proved to be efficient for the development of this project, so it can be applied in Internet of Things learning in educational institutions.

KEYWORDS: Internet of Things, Big Data, Monitoring.

1. INTRODUÇÃO

Esta proposta de pesquisa tecnológica está associada ao contexto Big Data, o qual tem provocado em diferentes segmentos industriais e, também, no ensino e pesquisa o interesse em explorar dados para processos decisórios (Gandomi & Haider, 2015). O interesse pelo lado industrial se deve em grande parte ao volume de dados gerados principalmente pelo uso de sensores no monitoramento de coisas, conceito chamado de Internet das Coisas ou simplesmente IoT (do inglês Internet of Things) (Chen, Mao & Liu, 2014). A razão no interesse de uso do histórico de dados para tomada de decisão industrial está associada a construção de modelos analíticos, principalmente baseados em algoritmos de inteligência artificial (IA), os quais possibilitam predizer situações antes das mesmas acontecerem, como é o caso da manutenção preditiva, assunto dentro do escopo da Indústria 4.0 (Lee, Kao & Yang, 2014; Wang, 2016).

Por outro lado, se tem observado o constante interesse da comunidade científica em pesquisas envolvendo Big Data e IoT. Estas são áreas que vem contribuindo significativamente ao ensino de Science, Technology, Engineering, e Mathematics (STEM) (Wu & Anderson, 2015; Glancy et al., 2017). Na literatura há relatos que o uso de Big Data e IoT como conteúdo de aula estimulam o interesse dos estudantes na aprendizagem (Marquez et al., 2016, Glancy et al., 2017; Oh et al., 2017; Otieno et al., 2017).

Estes relatos com viés de ensino e pesquisa motivaram a proposta deste trabalho, mas ao mesmo tempo, se transformaram em desafio no modo de como inserir Big Data e IoT em sala de aula e na pesquisa da graduação, considerando que os assuntos são de certa forma abstratos aos alunos, assim como grande parte da computação.

Portanto, de maneira objetiva, o problema de pesquisa deste trabalho está no desafio de como abordar Big Data e IoT de forma a estimular a imaginação do aluno para um despertar de interesse no ensino de STEM e, ao mesmo tempo, servir como contexto a aplicações do setor produtivo. Este problema emergiu do Projeto de Extensão “Aplicações de conceitos Big Data em problemas das industrias brasileiras” que tem como proposta a transferência destas tecnologias ao setor produtivo.

Justifica-se como proposta de Iniciação Tecnológica a necessidade em transformar as pesquisas desenvolvidas no laboratório referentes à Big Data e Métodos Analíticos em um ambiente de fácil assimilação aos estudantes de como pesquisas científicas podem ser aplicadas em problemas reais e de utilidade pública. Algumas ações já foram iniciadas neste sentido por meio da construção de um ambiente aos moldes de um ensaio de laboratório, com foco no desenvolvimento de ensino e pesquisa de estudantes de graduação em engenharia

ou computação com potencial de ser extensível ao ensino fundamental e também a pósgraduação.

Ao que cabe como desmembramento do Projeto Extensionista ao qual a pesquisa se vincula, justificado por resultados de Prova de Conceitos, a proposta consiste na elaboração de um laboratório didático científico baseando-se em um cenário de aquário. Trata-se de um elemento conhecido de todos, em que se sabe a priori todas as condições de contorno do ambiente (água, peixe, oxigênio, alimentação e etc) e, ao mesmo tempo, quando se coloca conceitos fundamentais de Big Data e IoT, o aluno rapidamente consegue fazer conexão com tal cenário. Portanto, o aquário serve como um contexto lúdico, no sentido que o assunto deixa de ser abstrato e passa a ser concreto, possibilitando ao professor passar claramente ao aluno o que se deseja realizar em uma sala de aula e, mais interessante ainda, da forma que o projeto foi concebido, como será detalhado adiante, estimula a criatividade do aluno em desenvolver uma pesquisa incremental.

O AcquaSmart, como o projeto foi batizado, não tem como pretensão automatizar um aquário, mas sim em criar um arcabouço em que se tenha diferentes sensores acoplados ao mesmo, gerando medidas para serem posteriormente traduzidas em informações estratégicas para tomada de decisão. Como exemplo de aplicação e trabalhos incrementados, o projeto iniciou com um sensor para medir a temperatura e um controlador para manter a temperatura constante. Como interface, o projeto tem painéis para visualização da temperatura em tempo real e análises estatísticas com o histórico de medidas e acionamento do aquecedor. O ambiente ainda possibilita baixar os dados históricos para fins de ensino e pesquisa (Sampaio, Correia e Silva, 2019). O trabalho já fomentou uma pesquisa de Iniciação Científica, a qual explorou a visualização de dados e a interface com o usuário com uso de realidade aumentada (Vieira, Silva e Martins, 2018). Uma combinação das duas pesquisas anteriores está sendo desenvolvida como Trabalho de Conclusão de Curso para validar o arcabouço a partir de medidas de pH da água e, por fim, os estudantes de graduação e pós-graduação que frequentam o laboratório desenvolveram um alimentador automático para o peixe, incluindo a prototipação e construção de um repositório de alimento e incremento na interface com o usuário para visualizar o horário que a ração é servida ao peixe. Em resumo, o AcquaSmart contempla a utilização de sensores para medir a temperatura, o pH e o horário que a ração é servida ao peixe e, ainda, tem dois controladores, o de aquecimento e alimentador. Como interface, o projeto tem painéis para visualização das medidas em tempo real e análises estatísticas do histórico de dados.

Ressalta-se, contudo, que os conceitos por de trás desta proposta servem a muitas outras áreas de aplicações que poderiam ter como base o AcquaSmart. Assim o uso deste

ambiente em exposições científicas, eventos acadêmicos ou mesmo feiras tecnológicas possibilita ao setor produtivo identificar o potencial da pesquisa em outros contextos, contribuindo assim para a transferência tecnológica ou, até mesmo, para um trabalho com viés inovativo.

1.1. Objetivo Geral e Específicos

O objetivo geral deste projeto é ampliar o escopo analítico do AcquaSmart a partir da implementação de métodos de predição de dados baseados em Inteligência Artificial, em específico Redes Neurais Artificiais.

Como objetivo específico, propõe-se a criação de cenários controlados como aumento de temperatura com níveis conhecidos de aquecimento, resfriamento da água, alteração controlada de pH e, também, simulações de falha do aquecedor e alimentador. Os cenários serão criados em um aquário sem a presença de um peixe, portanto, apenas um recipiente de água com os sensores e atuadores. Estes dados simulados serão usados neste trabalho para fins de validação das redes neurais, mas também disponibilizados a comunidade como um repositório público para fins de ensino e pesquisa em Big Data e IoT.

Por fim, dentro ainda dos objetivos específicos, será desenvolvido um módulo para que pessoas autorizadas possam coletar os dados do AcquaSmart em tempo real, o que é útil para diferentes contextos de ensino e pesquisa.

2. REFERENCIAL TEÓRICO

Segundo Chen, Mao e Liu (2014), Big Data é um conceito abstrato e deve ser definido quando se tem as seguintes características: geração e coleção de dados em grande volume, processamento dos dados devem ser realizados de forma rápida para que o resultado possa ser usado em tempo de operação e os tipos de dados são apresentados em diferentes formatos como estruturado, semi-estruturado e não estruturado. Estas características são tradicionalmente conhecidas como 3 V's (Zikopoulos et al., 2011; Chen, Mao e Liu, 2014). No entanto, outros autores acrescentam outras características como o valor agregado a um negócio, a partir da implementação de um Big Data e, ainda a incerteza dos dados disponíveis com qualidade fraca para a tomada de decisão, isto é, valor e veracidade, portanto, 5 V's (Chen, Mao e Liu, 2014). Embora ainda se encontrem outros autores acrescentando novos v's às características do Big Data mensuráveis são mesmo o volume, velocidade e variedade. O valor e outras características são subjetivas ao problema em análise e, portanto, não há consenso de definição.

A principal utilidade de Big Data é quando usado em tomada de decisões. Para possibilitar uma tomada de decisão baseada em evidências, as organizações precisam de processos eficientes para transformar grandes volumes de dados em insights significativos.

O processo geral de extrair insights do Big Data pode ser dividido em cinco estágios (Labrinidis & Jagadish, 2012; Gandomi e Haider, 2015), mostrados na Figura 1. Esses cinco estágios formam dois subprocessos: gerenciamento de dados e análise. O gerenciamento de dados envolve processos e tecnologias de suporte para aquisição e armazenamento de dados e para prepará-los e recuperá-los para análise. Análise, por outro lado, refere-se a técnicas usadas para analisar e adquirir inteligência de big data. Assim, a análise de big data pode ser vista como um subprocesso no processo geral de "extração de insight" a partir de grandes dados.

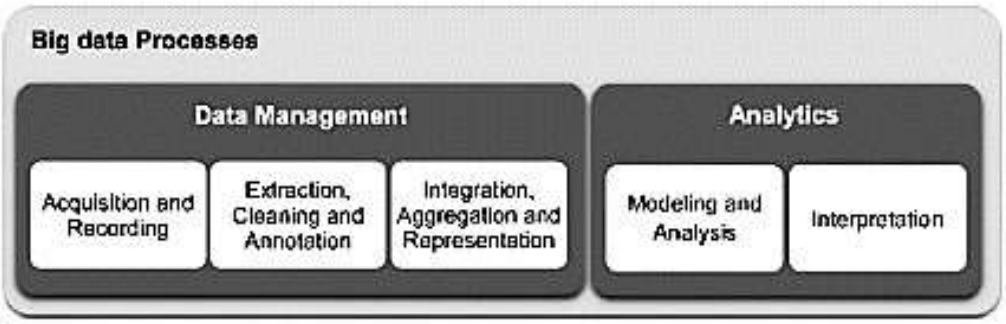

Figura 1: Processo para extração de insight de Big Data (Gandomi e Haider, 2015).

Na etapa de gerenciamento, a aquisição e registro é dependente do domínio de aplicação. No caso deste trabalho, a aquisição em via de regra é feita a partir de sensores. Entretanto, ainda há outras fontes de dados que se relacionam com os sensores como, por exemplo, anotações sobre situações de operações de funcionamento de um equipamento. Por outro lado, os estágios seguintes do gerenciamento são invariantes ao problema e devem ser considerados em grande parte de projetos Big Data, principalmente aos que estão relacionados a esta pesquisa.

Os problemas envolvidos no estágio de extração, limpeza e anotação, principalmente quando a fonte primária de dados é sensores, são relacionados principalmente a *noise*, *missing values* e *outlier*. Desta forma, conhecer a origem dos dados e o significado da média que está sendo coletada é de fundamental importância para saber o que está sendo armazenado e a qualidade dos dados que serão analisados. A integração, agregação e representação são tipos de pré-processamento que são feitos nos dados de modo a preparar para o módulo de análise (Han e Kamber, 2015; De Castro e Ferrari, 2017; Silva, Peres, Boscarioli, 2017).

2.1 Data Management

Como brevemente introduzido, quando se tem muitas fontes de dados é necessário fazer a integração das mesmas. Nesse processo, pela diversidade dos dados, é comum que os mesmos apresentem ruídos (noise data), inconsistências ou valores faltantes (missing values). Contudo, uma base de dados única não está dispensada de apresentar estes problemas (Garcia et al., 2015; Han et al., 2011; Acuña, 2011; Fayyad et al., 1996). Portanto, o pré-processamento que trata especificamente o cleaning da Figura 4.

A limpeza consiste basicamente em eliminar dados com ruídos (*noise data*), corrigir dados inconsistentes e manipular dados ausentes (*missing values*). Um dado ruidoso consiste na identificação de atributos com valores fora de um padrão esperado. As causas são diversas como, por exemplo, variação de uma medição em um equipamento, problema de medição de um humano entre outros. As soluções para tratamento de valores ruídos podem ser pela simples remoção do valor, caso de fato seja identificado como sendo um outlier, ou, por tratamento com uso de *binning* ou *clustering*. No entanto, a de se considerar que o surgimento de um valor diferente do normal, chamado convencionalmente de outlier, pode ser o resultado de uma medida nunca vista (Han e Kamber, 2015; Silva, Peres, Boscarioli, 2017).

A inconsistência de dados, por outro lado, consiste na presença de valores conflitantes em um mesmo atributo. É causada principalmente pela integração de diferentes bases de dados. Ou seja, se cada base de dados usa uma escala diferente para o mesmo atributo como, por exemplo, uma medida em quilowatt outra em megawatt, na integração os valores serão diferentes. Neste caso de inconsistência, a correção pode ser feita de maneira manual ou ainda considerando algum tipo de normalização, como apresentado adiante em Transformação de Dados. A solução para tratar a inconsistência de dados pode ser uma simples remoção ou a correção manual dos valores.

A ausência de dados, como o próprio nome sugere, ocorre quando um ou mais valores de atributos não existem. As causas podem ser diversas como, falha no preenchimento manual, não conhecimento do atributo, atributo com baixa importância entre outros. O problema pode ser resolvido de forma simples, removendo o atributo ou, se for um problema em outros atributos do mesmo exemplo, removendo todo o exemplar. E, também, o problema pode ser resolvido com técnicas mais elaboradas. Uma delas é calculando valores médios (mínimos ou máximos) do atributo e atribuí-lo a todos os valores ausentes.

A redundância de dados consiste quando dois atributos têm os mesmos valores ou são muito parecidos. Ela pode existir por diferentes motivos. Por exemplo, falta de informação adicional da base de dados (metadados) para se ter informação que um atributo é derivado de outro. Ainda há de considerar que a redundância pode existir entre exemplares da base.

A seleção permite obter uma base com representação reduzida em volume, sem redundância, mas com idêntica capacidade analítica. O resultado é a redução de dimensão em número de exemplares e, também, atributos. Tipicamente a redundância pode ser identificada com a utilização da análise de correlação, mas também pode ser feita manualmente ou então usando técnicas como, por exemplo, Análise de Componentes Principais ou PCA (do inglês, Principal Components Analisys). Na análise de correlação, o Coeficiente de Correlação de Pearson ou simplesmente ρ é um dos mais frequentes utilizados.

A transformação dos dados, como o nome sugere, permite a transformação do tipo de atributos. É possível a transformação de um valor numérico contínuo para discreto, um valor discreto para categórico ordinal, categórico nominal para discreto binário e categórico ordinal para discreto. E ainda, é possível nesta etapa do pré-processamento que atributos com valores em intervalos amplos sejam normalizados, de tal maneira que eles tenham mesma importância em um processo de mineração de dados. Na literatura há diferentes métodos de normalização como o z-score que transforma os valores dos atributos para que fiquem com média zero e desvio padrão igual a um e o método que muitos autores consideram como padrão que é o min-max.

Como discussão final sobre as etapas disponíveis na fase de Pré-Processamento, elas devem ser analisadas no aspecto de verificar a necessidade de utilização. Em caso todas elas se façam necessário, o natural é que à sequência preparação, seleção e transformação seja mantida. Ou seja, um processo iterativo. No entanto, não há restrição para que haja interação entre as etapas, isto é, que ao final de uma etapa não haja a necessidade de refazer a anterior. Em resumo, as etapas do Pré-Processamento se identificam com o processo do KDD, podendo ser iterativo e interativo.

2.2.Big Data Analytics

A análise pode ser basicamente subdividida em análise exploratória de dados, que usam recursos de estatística descritiva ou métodos de agregação em bases de dados ou por técnicas que envolvem tarefas de mineração de dados.

Abordagem baseada em Análise Exploratória são aquelas que usam basicamente a linguagem SQL (do inglês, *Structured Query Language*) para auxiliar na montagem de uma base de dados a partir de consultas diversas e em diferentes fontes de dados e, também, na

sumarização dos dados com uso de funções agregadas que manipulam os valores através de operações como soma, média, mínimo, máximo de valores, contagens e etc, as quais também são importantes para a construção de relatórios consolidados (Han e Kamber, 2015; Silva, Peres, Boscarioli, 2017)..

Exemplo típico desta abordagem analítica é a partir da construção de modelos de dados multidimensionais. Este modelo baseia-se em um armazém de dados (*Data Warehouse*), cujos valores são extraídos dos dados transacionais OLTP (do inglês, *On-Line Transaction Processing*). Estes são dados usados exclusivamente para análise e, portanto, não devem ser atualizados (não-voláteis). A exploração desta base de análise pode ser feita através de ferramenta de processamento analítico on-line ou OLAP (do inglês, *Online Analytical Processing*). A aplicação dessa ferramenta permite que se explorem diferentes perspectivas de uma base de dados e, também, que se alterem os valores em prol de objetivos analíticos.

Portanto, são consultas feitas às bases de dados com perguntas estabelecidas ou, então, modelagens de dados relacionando tabelas definidas pelos gestores que são preparadas por meio de um processo de extrair, transformar e carregar dados, popularmente conhecido por ETL (do inglês, *Extract, Transform, Load*), a fim de se ter um repositório analítico para descobertas desejáveis.

Diferentemente da abordagem acima exposta, a baseada em mineração de dados consiste em um processo automático ou semi-automático de explorar analiticamente grandes bases de dados com a finalidade de fazer descobertas que não são de conhecimento óbvio do especialista, mas que o seu resultado serve de insumo para que o mesmo possa usá-lo em processos de tomadas de decisões. A Mineração de dados é componente principal de um processo de descoberta de conhecimento em bases de dados (KDD, *Knowledge Discovery in Databases*). O processo de KDD inicia com a organização das bases de dados, que contêm registros da área de interesse, a partir das quais se deseja descobrir algum tipo de informação relevante para, por exemplo, ajudar na tomada de decisão. Na fase seguinte, pré-processamento, os dados são preparados em um repositório único e normalmente precisam ser tratados para eliminar exemplares repetidos e ou valores discrepantes. Ainda no pré-processamento, a seleção de dados permite que se escolha apenas os exemplos ou atributos relevantes para a mineração de dados. Adicionalmente ainda nesta fase do processo, os valores dos atributos selecionados podem ser normalizados para, por exemplo, terem o mesmo domínio de valores e, consequentemente, sejam considerados com mesma importância no processo de Mineração de Dados. Com os dados processados se inicia a tarefa de Mineração de Dados propriamente dita, envolvendo técnicas de predição, clusterização ou

associação de dados. Estas técnicas utilizam algoritmos de aprendizagem de máquina e inteligência artificial. No caso específico da predição, os algoritmos são capazes de ajustar seus parâmetros livres a partir do histórico de dados. A predição pode ser categórica, quando por exemplo deseja-se saber o tipo de falha que houve em um equipamento ou pode ser numérica, com a inferência do melhor período para a manutenção de um equipamento. Por fim, os resultados apresentados pela mineração de dados são analisados e validados. Todo o processo de KDD é iterativo, feito passo a passo, e interativo, a sequência de cada fase pode ser alterada ou repetida (Fayyad, Piatetsky-Shapiro e Smyth, 1996).

4. METODOLOGIA

Para que os objetivos do trabalho fossem alcançados estabeleceu-se que inicialmente fossem realizados estudos abrangendo, essencialmente:

a) Mineração de Dados: estudo dos capítulos introdutórios das seguintes referencias que permitirão entendimento do contexto de análise exploratória e todo processo de descoberta de conhecimento por meio de dados, ampliando os já iniciados e apresentados acima: (Han e Kamber, 2015; Silva, Peres, Boscarioli, 2017);

b) Redes Neurais: estudo das seguintes referências no que diz respeito a introdução do assunto e das arquiteturas de redes Multi-Layer Perceptron (MLP) e Multi-Layer Perceptron e Long Short Term Memory Networks (LSTM): Haykin, 2010; Aggarwal, 2018);

c) Literatura: As referências acima permitem se ter uma base fundamental para entendimento dos conceitos envolvidos na pesquisa. Entretanto, a revisão bibliográfica sobre o assunto no portal Capes será feita em busca de trabalhos relacionados.

Como procedimento metodológico técnico do projeto, as seguintes etapas serão desenvolvidas:

a) Definição dos seguintes cenários para geração de dados: situações de esfriamento da água de forma leve, média e abrupta;

b) Geração de cenários para modelagem preditiva: criação de situações que indiquem falha do aquecedor da água;

c) Implementação do módulo preditivo na interface do AcquaSmart: na interface das medidas em tempo real serão incorporadas medidas de inferência. Assim, na interface de histórico, os erros entre as medidas reais e as previstas serão apresentadas. Assim, caso a diferença entre as medidas seja grande terá a opção de refazer o treinamento da rede

neural; d) Opção do modelo preditivo: no módulo preditivo será implementado a opção ao usuário selecionar a rede neural que quer usar para a predição, possibilitando uma análise comparativa;

e) Criação de um módulo de compartilhamento: no histórico de dados será possível baixar os dados reais ou os simulados para uso em fins didáticos e de pesquisa. Além disso, será proposto uma API para que os dados reais possam ser coletados.

5. RESULTADO E DISCUSSÃO

Todo ambiente computacional envolvido no projeto utiliza o estado da arte de pesquisa e tecnologia em Big Data e IoT. A maneira como o projeto foi desenvolvido pode ser útil para contribuir com a inovação em termos de produto como, por exemplo, no monitoramento de água, ar e outros. No entanto, qualquer outra demanda que envolvam Internet das Coisas, Smart City, Big Data, Métodos Analíticos e Inteligência Artificial são potenciais de aplicação deste projeto. Por exemplo, monitorar a qualidade de água de um rio e criar um modelo de inferência para saber com antecedência se após alguns anos a água estará adequada para consumo.

Com este arcabouço, alunos de graduação e pós que queiram desenvolver pesquisas no laboratório podem, por exemplo, acoplar outros tipos de sensores como para medir o nível de oxigênio da água, criação de modelos de inferência baseados em Inteligência Artificial, métodos de visualização de dados, segurança de informação, arquiteturas de Big Data e etc.

A Figura 2 apresenta a arquitetura básica que possui 3 componentes: o controlador, o servidor de dados e a aplicação web. Os dados coletados pelo controlador são apenas armazenados em um servidor de dados (IoT) e apresentados na aplicação web. Esta interação, entretanto, pode acontecer em qualquer sentido, permitindo o controlador receber comandos da aplicação web a partir da leitura de valores no servidor de dados. O controlador é responsável pela coleta de dados no ambiente por meio de sensores, ações de controle e transmissão dos dados ao servidor. O servidor, por sua vez, armazena de forma estruturada os dados servindo de integrador entre o controlador e a aplicação web. A aplicação web, por sua vez, permite aos usuários visualizar o comportamento do ambiente e acessar os dados coletados para conduzir estudos futuros.

Referente a aplicação, as Figuras 3 e 4 apresentam as interfaces desenvolvidas, sendo de forma respectiva, a visualização das medidas em tempo real e dos últimos 30 minutos e o histórico de dados com estatísticas descritivas realizadas para períodos específicos de análise.

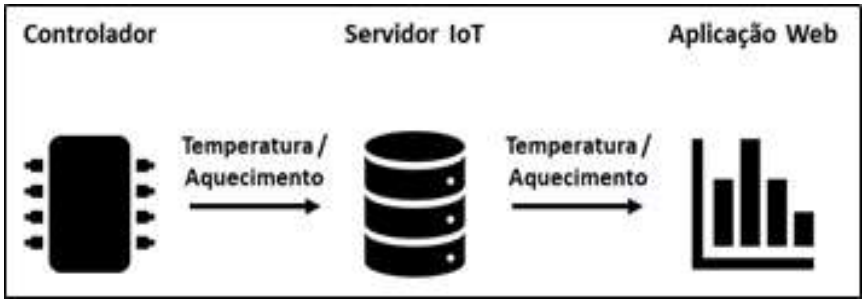

Fig. 2: Arquitetura do projeto

Neste projeto está sendo considerado que o aquário hospedará um único peixe da espécie Betta. Isto é relevante para que possa ser possível determinar os índices de temperatura ideal da água para sobrevivência deste tipo de peixe que é entre 22ºC e 28ºC. Por isto, na Figura 2 há duas linhas contínuas nestas temperaturas.

O ambiente externo ao aquário sofre interferência de temperatura de ar condicionado e, também da variação da temperatura ambiente conforme mudança de estações do ano ou número de pessoas no local. Para manter a temperatura nos padrões ideais para o peixe beta, utilizou-se como controlador o Arduino Uno. O primeiro motivo da escolha deste microcontrolador foi a facilidade de aprendizado e operação de suas funcionalidades, sendo ideal para utilização com estudantes do ensino médio a pós-graduação. O segundo motivo foi o baixo preço.

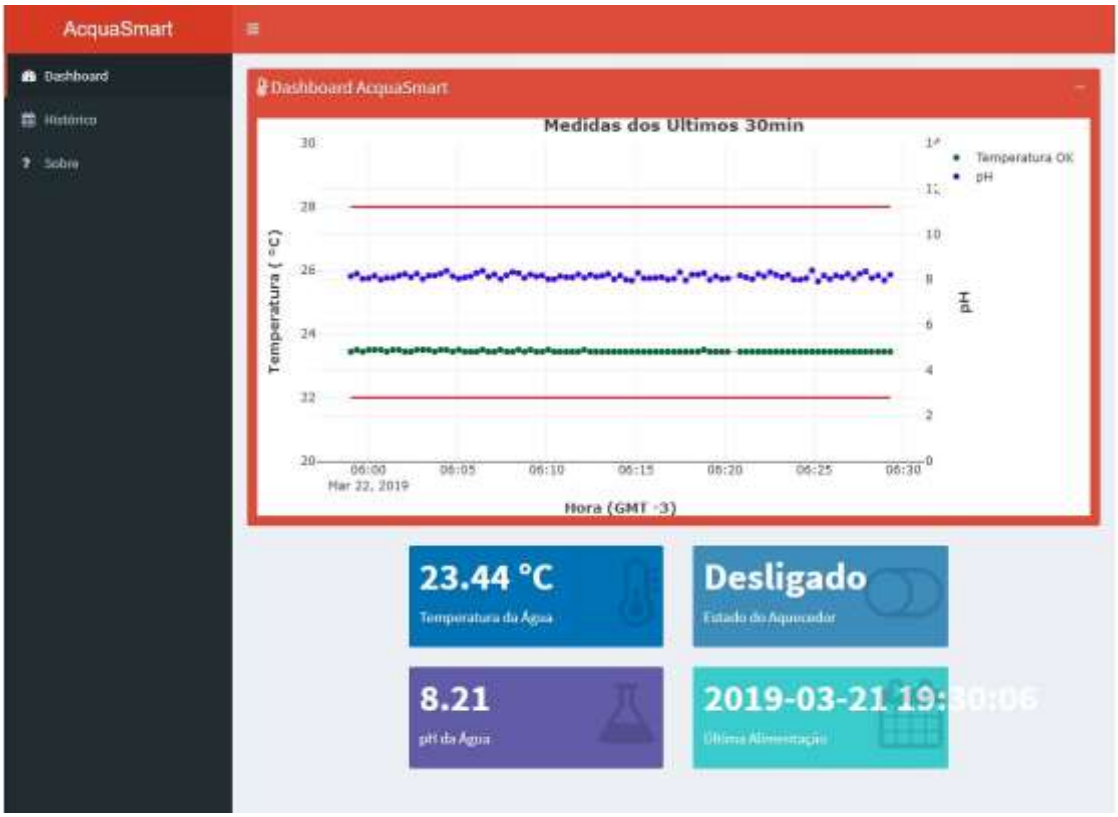

Figura 3: Dashboard com as medidas em tempo real.

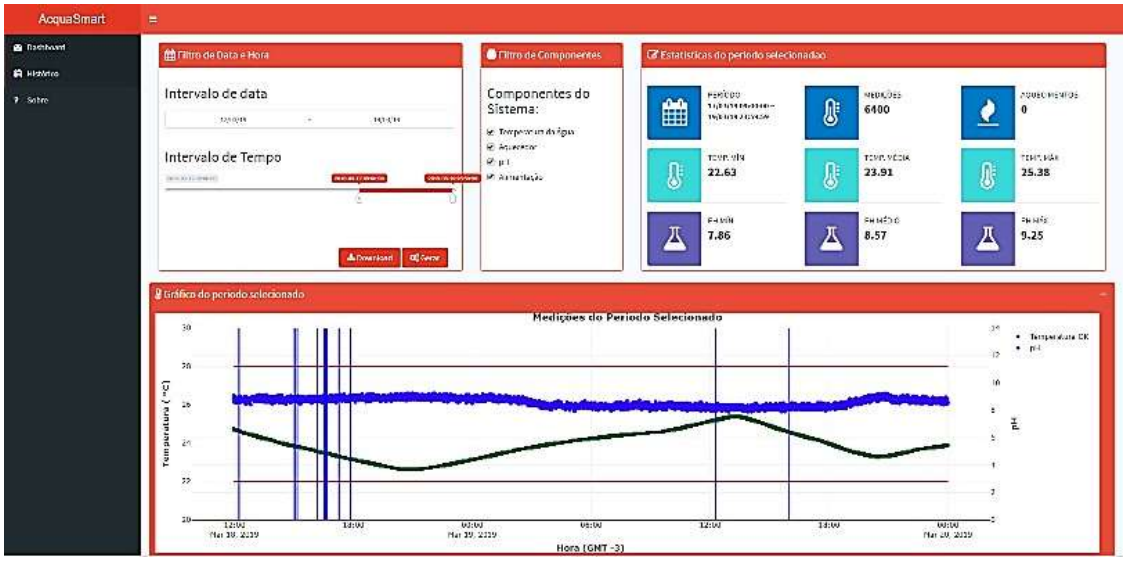

Figura 4: Dashboard com o histórico de medidas em tempo real.

De forma a evitar que o comportamento da temperatura da água se mantenha de maneira mais estável, estudo de *forecasting* foram realizados de forma a avaliar a possibilidade de projeção de valores futuros com uso de uma rede neural do tipo Múltiplas Camadas (MLP) tendo como entrada medidas de uma semana e saída o dia seguinte. O conjunto de dados normalizado para estes experimentos está ilustrado na Figura 5. Nestas medidas de temperatura foram introduzidos ruídos como valores faltantes a fim de analisar o impacto na modelagem da projeção.

O treinamento do modelo foi feito com quase todos os dados disponíveis (Jan de 2018 a Jan 2020), sendo que o objetivo era avaliar a projeção para o mês de fevereiro de 2020. Para avaliar o desempenho do modelo, projetou-se o resultado pelo modelo junto a série de medidas de temperatura e, como se observa da Figura 6, há sobreposição de valores, que indica boa aderência entre dados reais e dados projetados (*forecasting*) e que tem a medida de desempenho feita pela raiz quadrada do erro médio quadrático (RMSE) de 0.0023.

Para fins de análise de desempenho, usou-se o modelo para projetar o mês de fevereiro, sendo que esse período não foi usado no treinamento. O resultado da projeção destas medidas está ilustrado na Figura 7, cujo RMSE foi de 0.0043.

Finalmente projetou-se a medida para o mês de março cujas medições não continham disponíveis nem para treinamento e nem para teste, como forma de mostrar o uso operacional do projeto.

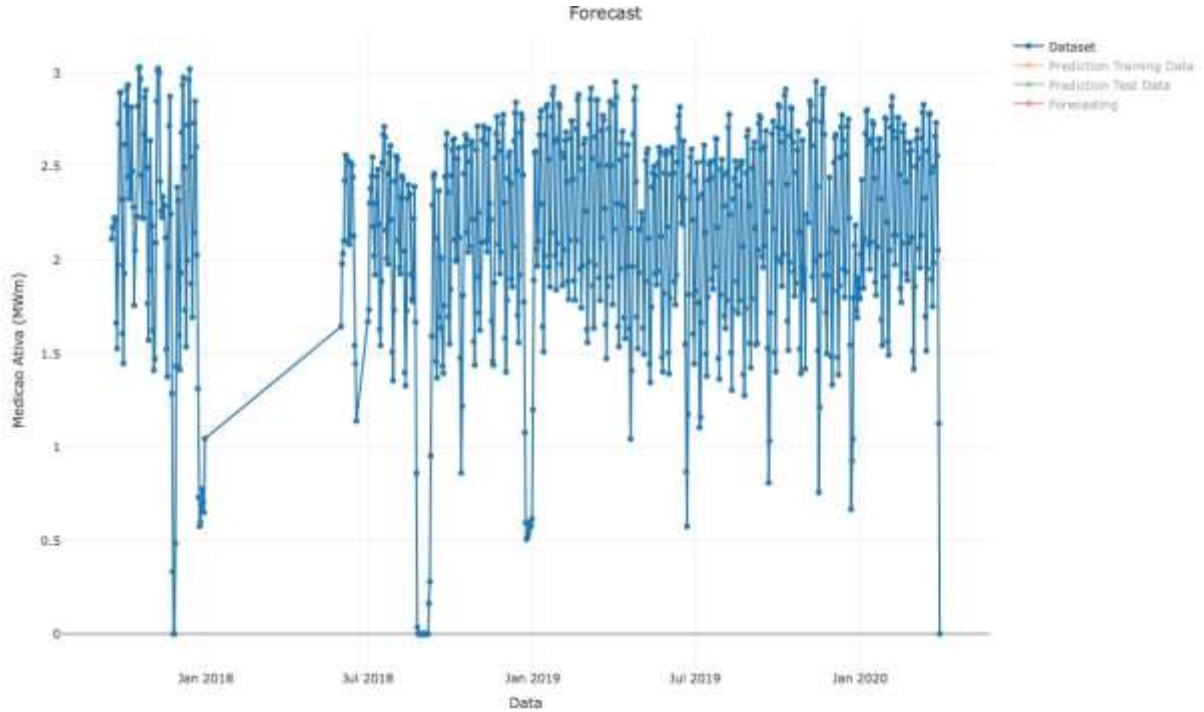

Figura 5: Série temporal usada no treinamento de uma rede neural

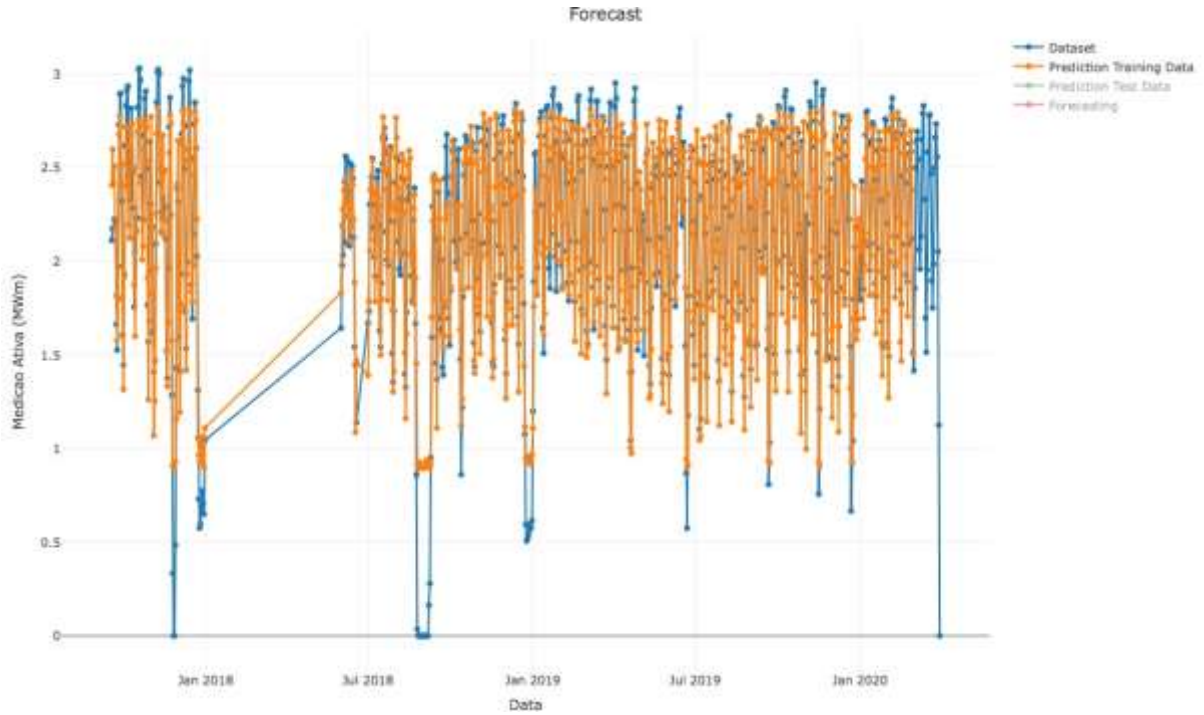

Figura 6: Sobreposição da série usada no treinamento e a projetada pelo modelo.

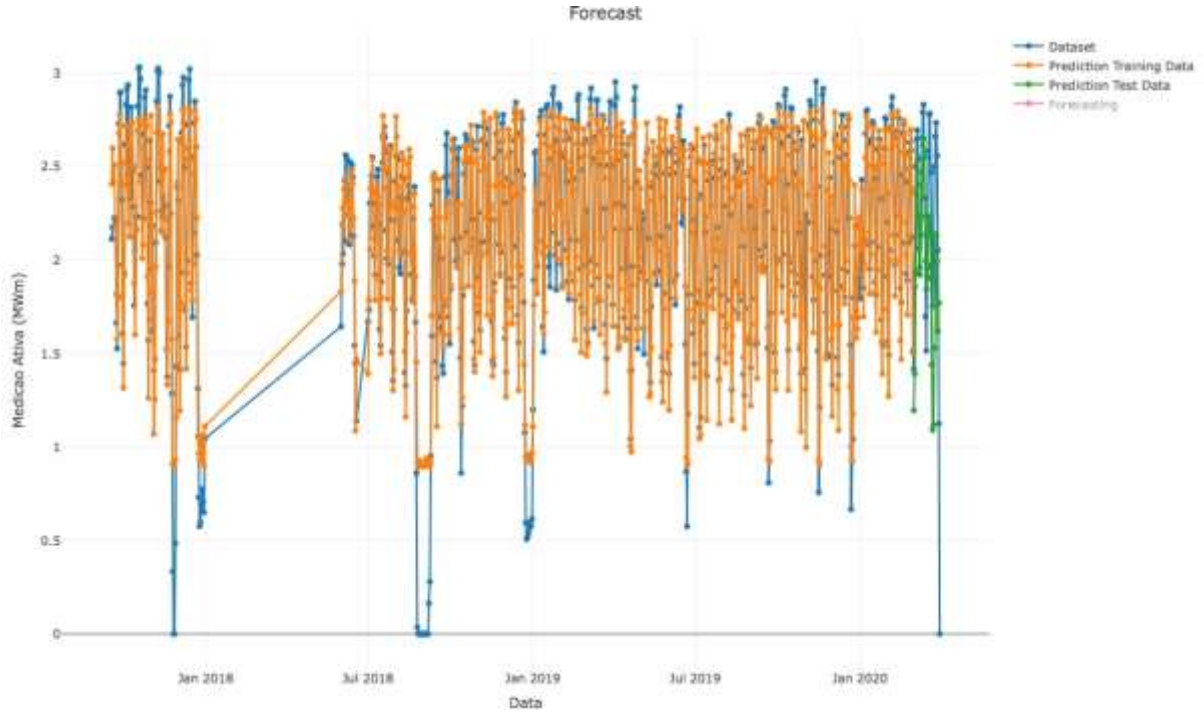

Figura 7: Sobreposição da série usada no teste e a projetada pelo modelo.

Figura 8: Projeção de medidas para o mês de março de 2020 (forecasting).

6. CONSIDERAÇÕES FINAIS

Pode-se concluir com os resultados demonstrados anteriormente, que a técnica estudada apresentou resultados satisfatórios para o objetivo proposto. A plataforma é flexível para fins de pesquisa incremental e fins didáticos.

Os experimentos de projeção (forecasting) ainda precisam de serem mais bem explorados, uma vez que a pesquisa foi fortemente impactada pela pandemia com o fechamento do laboratório e queima dos sensores que monitoravam o aquário.

Ainda assim, com os dados que tinham disponíveis foi capaz de validar a possibilidade de projeção, analisado pelo RMSE de treinamento e teste que se mostraram semelhantes, levando a concluir que o modelo de RNA utilizado é capaz de projetar medidas futuras.

Como trabalhos futuros pretende-se acoplar o módulo de projeção (*forecasting*) na plataforma e analisar os resultados de projeção no sentido de descobrir para este tipo de aplicação quantos dias ou meses a projeção é capaz de manter um desempenho confiável. Ainda como trabalho futuro pretende-se analisar a depreciação do modelo frente aos dados históricos e necessidade de retreinamento do modelo.

7. REFERÊNCIAS

Acuña, E. Preprocessing in Data Mining. In International Encyclopedia of Statistical Science. Springer, Berlin, Heidelberg, pp. 1083-1085, 2011.

Aggarwal, C. C. Neural networks and deep learning. Berlin, Germany. Springer, 2018.

Chen, M., Mao, S., & Liu, Y. Big data: A survey. Mobile networks and applications, 19(2), 171209, 2014.

De Castro, L. N. & Ferrari, D. (2017). Introdução a mineração de dados. Editora Saraiva.

Fayyad, U., Piatetsky-Shapiro, G., & Smyth, P. From data mining to knowledge discovery in databases. AI magazine, 17(3), 37, 1996.

Gandomi, A., & Haider, M. Beyond the hype: Big data concepts, methods, and analytics. International Journal of Information Management, 35(2), 137-144, 2015.

García, S., Luengo, J., & Herrera, F. Data preprocessing in data mining. Switzerland: Springer International Publishing, pp. 195-243, 2015.

Glancy, A. W., Moore, T. J., Guzey, S., & Smith, K. A. (2017). Students' Successes and Challenges Applying Data Analysis and Measurement Skills in a Fifth-Grade Integrated STEM Unit. Journal of Pre-College Engineering Education Research (J-PEER), 7(1), 5.

Han, J., Pei, J., & Kamber, M. Data mining: concepts and techniques. 1a. ed., Elsevier, 2015.

HAYKIN, Simon. Neural networks: a comprehensive foundation, 1999. Mc Millan, New Jersey, p. 1-24, 2010.

Labrinidis, A., & Jagadish, H. V. Challenges and opportunities with big data. Proceedings of the VLDB Endowment, 5(12), 2032–2033, 2012.

Lee, J., Kao, H. A., & Yang, S. Service innovation and smart analytics for industry 4.0 and big data environment. Procedia Cirp, 16, 3-8, 2014.

Lecun, Y.; Bengio, Y.; Hinton, G. Deep learning. Nature, v. 521, n. 7553, p. 436, 2015.

Marquez, J., Villanueva, J., Solarte, Z., & Garcia, A. IoT in Education: Integration of Objects with Virtual Academic Communities. In New Advances in Information Systems and Technologies, pp. 201-212, 2016.

Oh, T. T., Chung, S., Lunt, B., McMahon, R., & Rutherfoord, R. The Roles of IT Education in IoT and Data Analytics. In Proceedings of the 18th Annual Conference on Information Technology Education, pp. 39-40, 2017.

Otieno, W., Cook, M., & Campbell-Kyureghyan, N. Novel approach to bridge the gaps of industrial and manufacturing engineering education: A case study of the connected enterprise concepts. In 2017 IEEE Frontiers in Education Conference (FIE), pp. 1-5, 2017.

Sampaio, G. S.; Correia, G. M.; Silva, L. A. AcquaSmart: An Environment Big Data Analytics and Internet of Things to Education and Research. International Journal for Innovation Education and Research, v. 7, n. 1, p. 93-104, 2019.

Silva, L. A., Peres, S. M., & Boscarioli, C. Introdução à mineração de dados: com aplicações em R. 1ª. Edição, Elsevier Brasil, 2017.

Vieira, D. D. ; Silva, L. A. ; Martins, V. F. . Uso de Realidade Aumentada para Visualização e Interação com Dados de Sensores. In: Israel Florentino dos Santos; Paulo Nazareno Maia Sampaio. (Org.). I Workshop Latino-Americano De Trabalhos Em Andamento Em Computação (Wlatac). 11ed.São Paulo: Sociedade Brasileira de Computação, 2018, v. 1, p. 4-140.

Zikopoulos, P., & Eaton, C. Understanding big data: Analytics for enterprise class hadoop and streaming data. McGraw-Hill Osborne Media, 2011.

Wang, K. Intelligent predictive maintenance (IPdM) system–Industry 4.0 scenario. WIT Transactions on Engineering Sciences, 113, 259-268, 2016.

Wu, Y. T., & Anderson, O. R. Technology-enhanced stem (science, technology, engineering, and mathematics) education, 2015.

Contatos: abramgrossmann2016@gmail.com e leandroaugusto.silva@mackenzie.br