

Uso de Aprendizagem de Máquina em Dados Agrometeorológicos

Jorge S. Farias^{1,2}, Elthon Oliveira¹, Rodolfo Cavalcante¹

¹Universidade Federal de Alagoas (UFAL) - Campus Arapiraca
Avenida Manoel Severino Barbosa, s/n - Bom Sucesso, 57309-005 – Arapiraca – AL

²Bolsista PIBITI Edital 2019-2020 UFAL/CNPq/FAPEAL.

{jorge.farias, elthon, rodolfo.cavalcante}@arapiraca.ufal.br

Resumo. *Este artigo apresenta uma evolução de uma ferramenta de apoio à análise de dados agrometeorológicos. Esta ferramenta auxilia pesquisadores e agricultores familiares em seus trabalhos. A evolução se refere ao desenvolvimento de um módulo predictor de dados que são captados por uma estação agrometeorológica, que minimiza um problema de falta de dados. Este módulo prediz dados ainda não coletados a partir de dados já coletados. Para isso, é utilizada uma técnica de aprendizagem de máquina. Uma nova arquitetura e uma análise dos resultados obtidos são apresentados.*

Abstract. *This article presents an evolution of a tool to support the analysis of agrometeorological data. This tool helps researchers and family farmers in their work. The evolution refers to the development of a predictor module of data that is captured by the station, which minimizes a problem of lack of data. This module predicts data not yet collected from data already collected. For this, a machine learning technique is used. A new architecture and an analysis of the results obtained are presented.*

1. Introdução

Estações agrometeorológicas automáticas têm por finalidade monitorar as condições meteorológicas vigentes que permitem quantificar diversos dados físicos do ambiente, a saber: radiação solar, precipitação pluviométrica, velocidade do vento, direção do vento e pressão atmosférica. Estes dados são utilizados no manejo da água para irrigação e para auxiliar na tomada de decisão nos sistemas agropecuários animal e vegetal. Estes dados servem para dar suporte à pesquisa científica em áreas como monitoramentos ambiental e agrícola, para atender o interesse público para fins de planejamento agrícola de governos municipais e estaduais, e para gerar informações diárias sobre os elementos meteorológicos de uso comum pela sociedade.

O território brasileiro, em virtude de sua larga extensão, abrange uma ampla diversidade climática e muitas particularidades regionais. Neste sentido, segundo a [Seplag/AL 2015], a microrregião da cidade de Arapiraca-AL possui a maior densidade da agricultura familiar no Brasil. A região onde a cidade de Arapiraca está inserida é o semiárido brasileiro que é uma região delimitada pela Superintendência de Desenvolvimento do Nordeste (SUDENE) considerando condições climáticas dominantes de semiaridez, em especial a precipitação pluviométrica. Em Alagoas foram identificados no ano

de 2006, 123.331 estabelecimentos agrícolas [Veras 2011], sendo 90,6% dos estabelecimentos deste total pertencentes a agricultura familiar.

A agricultura familiar ainda é o principal meio de sobrevivência de uma parcela considerável população brasileira. Neste sentido, os elementos meteorológicos afetam não só os processos metabólicos das plantas, diretamente relacionado à produção vegetal, como também as mais diversas atividades no campo [Seplag/AL 2017].

Há no *Campus* Arapiraca (UFAL) uma estação automática da marca *Campbell Scientific* financiada com recursos FINEP e que está sob responsabilidade técnica de pesquisadores do grupo de pesquisa CRAD – Recuperação de Áreas Degradadas. Ela possui torre de 10 m com 18 diferentes sensores instalados e que, desde o final do ano de 2012, vem coletando dados a cada 10 minutos, em média, de dez variáveis distintas. Os preços para esse equipamento variam de acordo com a quantidade de sensores e equipamentos opcionais incluídos, e as especificações deles [Dousseau 2009].

Entretanto, devido à frequência de leitura e à quantidade de variáveis lidas, a quantidade de dados coletados dificulta a leitura/interpretação para uso nas atividades práticas e de cunho técnico-científicas. Pois, nem sempre os usuários querem visualizar todos os dados presentes na base de dados, e poder fazer essa visualização através de gráficos torna mais fácil o estudo dos dados. Esse problema foi resolvido anteriormente com a primeira versão da ferramenta.

Além disso, devido ao processo manual de coleta dos dados e à falta de um profissional dedicado a esta tarefa, frequentemente a base de dados não contém dados atuais. Há, no melhor caso, dados referentes ao dia anterior. Para contornar este problema, foi pensado em fazer uso de técnicas de aprendizagem de máquina para predição dos dados ausentes no banco de dados. Com o uso de tais técnicas, além de possibilitar a predição dos dados referentes à semana anterior - que não foram armazenados na base, por exemplo - a técnica possibilita a predição climatológica de dias ou semanas no futuro.

Os sistemas de aprendizagem de máquina podem usar dados históricos para prever os dados de forma muito mais precisa e rápida do que os seres humanos são capazes [DSA 2018]. Os dados previstos podem ser usados temporariamente enquanto os dados reais obtidos pela estação não são coletados.

Este artigo apresenta a implementação de um módulo previsor de dados em uma ferramenta existente, que tem como objetivo principal a geração de gráficos a partir dos dados obtidos pela estação. Este módulo previsor representa uma adição significativa ao trabalho já desenvolvido, aumentando a funcionalidade da aplicação.

2. Trabalhos Relacionados

No contexto da análise de dados meteorológicos, algumas ferramentas semelhantes à apresentada neste trabalho são discutidas a seguir.

CLIMAP – Aplicativo para análise de dados climáticos - É uma ferramenta computacional para análise de dados meteorológicos diários [Salvador 2017]. As variáveis meteorológicas analisadas são três fixas: chuva, temperatura máxima e temperatura mínima; e uma quarta variável escolhida pelo usuário. Nesse aplicativo são utilizadas técnicas estatísticas para análise de variação e tendência nas séries de chuva e temperatura. Diversos índices de extremos também são disponibilizados. A ferramenta disponibiliza

recursos para análise de correlação entre índices climáticos e as séries de chuva e temperatura. O aplicativo é de fácil instalação e uso, com uma interface gráfica simples e objetiva.

Ferramenta para visualização de dados meteorológicos - Desenvolvido na UFSC, uma ferramenta para auxiliar a análise de eventos meteorológicos [Matheus 2011], através da visualização de campos 3D de dados obtidos a partir do modelo ETA (O modelo de área limitada) gerado pelo INPE. A ferramenta serve para estudo de casos por especialistas, bem como para a apresentação de dados meteorológicos para o público leigo. A interface criada possibilita a interação do usuário com uma grande quantidade de dados em tempo real, permitindo ações de rotação, translação e zoom da câmera, bem como outras funções para a inspeção dos dados.

Previsão de Temperatura Atmosférica usando Suporte máquinas vetoriais - É uma aplicação de Máquinas vetoriais de suporte (SVMs) para previsão do tempo [Y. Radhika 2009]. Dados da série temporal de temperatura máxima diária em um local é analisado para prever a temperatura máxima do dia seguinte, naquele local, com base no máximo diário.

PULSE-Brasil - Uma ferramenta capaz de fazer análises integradas de dados sobre clima, ambiente e saúde pública e de fornecer cenários climáticos futuros [FAPESP 2014] para orientar a tomada de decisão relacionada a eventos extremos já está disponível gratuitamente na internet.

Previsão meteorológica e climática orientada por dados: Aproximando um modelo de circulação geral simples Com Aprendizado Profundo - Este trabalho tem como objetivo simular a dinâmica de um simples modelo de circulação geral com uma rede neural profunda [Scher 2018]. Depois de ser treinada no modelo, a rede pode prever o estado completo do modelo vários passos à frente — o que conceitualmente está fazendo previsões meteorológicas no mundo modelo.

Um sistema web para a consulta de dados meteorológicos como ferramenta de apoio ao manejo de irrigação no estado do Piauí - Sistemas de consulta a dados meteorológicos promovem uma utilização adequada dos recursos hídricos [Silva et al. 2008]. Um sistema web e acesso dinâmico à informações climáticas. O principal recurso do sistema é a geração automática de planilha de dados a partir do resultado da consulta ao banco de dados. O sistema web desenvolvido dinamizou a forma como os dados climáticos são manipulados e mostrou-se efetivo na recuperação de informações a respeito das variáveis climáticas, evidenciando sua aplicabilidade para subsidiar o manejo de irrigação das culturas no Estado do Piauí.

Considerações sobre os trabalhos relacionados

A principal diferença da ferramenta apresentada neste artigo para as supracitadas, excetuando-se três, é o uso das técnicas de aprendizagem de máquina. A maioria das ferramentas existentes apenas serve para a análise dos dados. Tais ferramentas não oferecem predição de dados. No entanto a *PULSE*, apesar de fornecer tal predição, limita o intervalo de tempo previsto.

A ferramenta aqui apresentada permite ao usuário definir o intervalo de tempo de maneira livre. Porém no contexto de previsões climatológicas há o risco de uma queda

na precisão dos valores preditos quando se trabalha com grandes intervalos de tempo. A *Previsão de Temperatura Atmosférica usando Suporte máquinas vetoriais* apenas oferece a previsão para a temperatura máxima diária. As previsões meteorológica e climática orientadas por dados apenas oferecem o modelo e não uma ferramenta executando propriamente dita.

3. Materiais e Métodos

3.1. Materiais

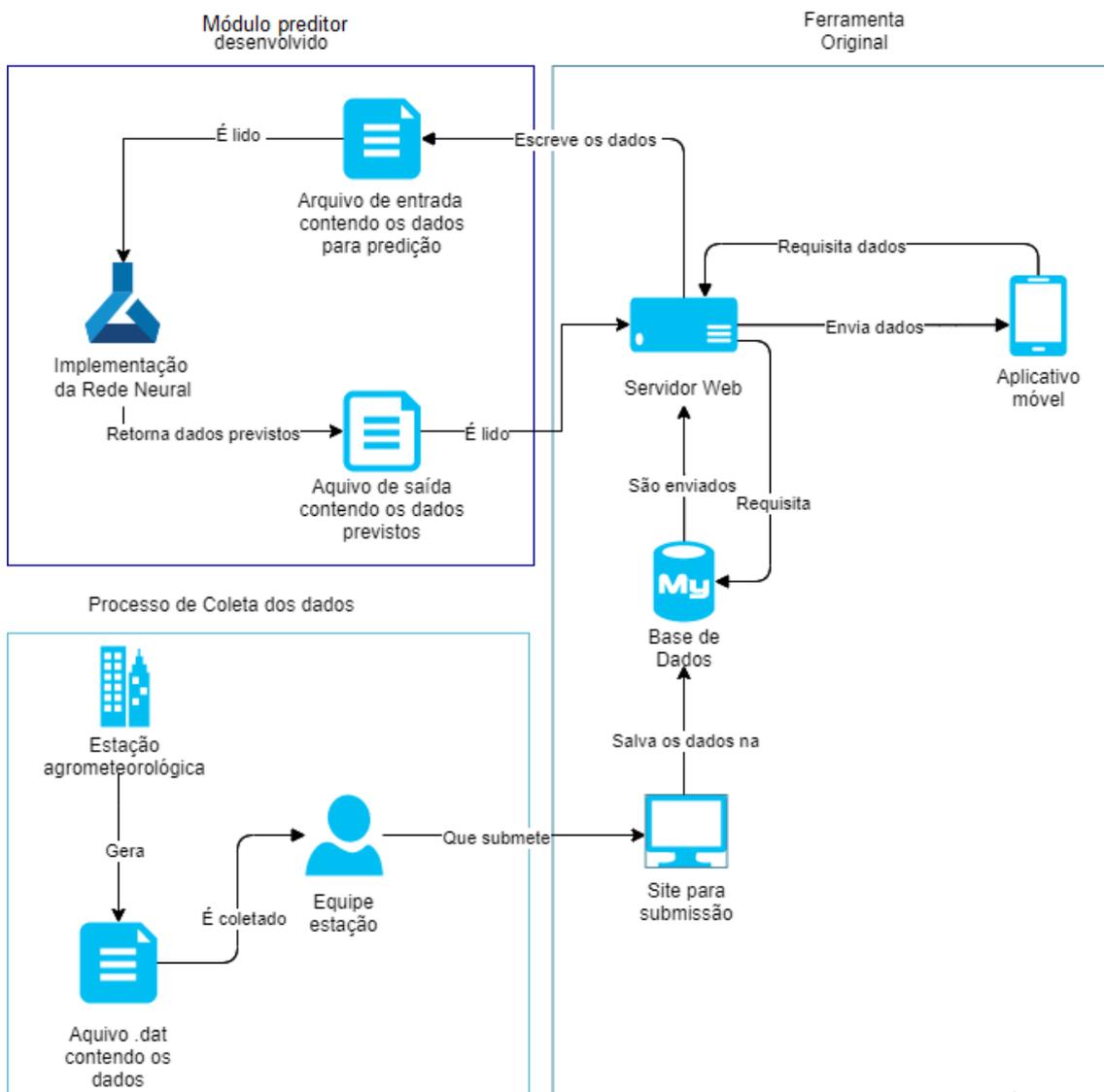


Figura 1. Diagrama de funcionamento da ferramenta.

Fonte: Elaborada pelo autor

Na Figura 1 é apresentado o processo de funcionamento da ferramenta. As setas na figura representam o fluxo de execução de coleta de dados e do funcionamento da ferramenta. Começando com o processo de alimentação de dados a estação gera os arquivos *.dat*. Os arquivos são submetidos em uma ferramenta criada com Thymeleaf 4 que é um

mecanismo para servidor java para construir ambientes web [thymeleaf 2018], que por sua vez envia os dados para serem processados e armazenados na base de dados SQL. No processo de execução da aplicação, primeiro o usuário acessa o aplicativo móvel, este aplicativo desenvolvido em Flutter, que é um framework para desenvolvimento de aplicativos mobile, recupera as informações depois de processadas. Os dados são recuperados por meio de requisições HTTP e servirão para implementação das técnicas de aprendizagem de máquina. A escolha do Flutter para o desenvolvimento do aplicativo se deu pelo fato de ser um SDK de código aberto criado pela Google [Flutter 2017], que vem crescendo no contexto de aplicações multiplataforma para Android e iOS desde o seu lançamento em maio de 2017.

A partir do momento em que a requisição do cliente é realizada, ela é recebida em um servidor implementado em Java utilizando o framework Spring [Spring 2020], que é uma ferramenta para criação de soluções da web.

No processo de predição, o algoritmo de aprendizagem de máquina utiliza redes neurais e é implementado em Python 3 com a biblioteca Scikit-learn [Scikit 2019], que é uma biblioteca de aprendizagem de máquina de código aberto que suporta aprendizados supervisionado e não supervisionado.

Algumas das tecnologias foram utilizadas por que já estavam sendo empregadas na versão anterior da ferramenta, sendo essas Flutter para desenvolvimento do aplicativo, Spring e Thymeleaf para o desenvolvimento do servidor. Já para o módulo previsor foi adotada a linguagem Python. Pois, essa linguagem apresenta uma ampla gama de bibliotecas para aprendizagem de máquina e é uma linguagem simples. Entende-se que quanto menos o desenvolvedor precisar se preocupar com o próprio código, mais foco e trabalho poderão ser colocados na busca de soluções.

Para definir a biblioteca de aprendizagem de máquina que seria usada, foram estudadas algumas opções, tais como: Tensor Flow, Keras e Scikit-learn. Esta última tem uma função específica para predição dos dados. Por isso, ela foi adotada para o desenvolvimento do módulo previsor.

3.2. Métodos

Para o desenvolvimento do módulo previsor e ajustes na ferramenta, utilizou-se de *metodologia ágil*, que é uma resposta à gestão de projetos tradicional e tem a função de aprimorar o processo de desenvolvimento de um produto ou serviço [dos Santos Soares 2009]. O objetivo desta metodologia é fazer entregas de código com rapidez e com maior frequência, conforme surgem as demandas do cliente.

Sobre o processo de aprendizagem, primeiramente é preciso fazer o pré-processamento dos dados coletados que é feito no momento em que o arquivo *.dat* é submetido no servidor. Assim, os valores presentes na base de dados já estão pré-processados para manter um padrão no *dataset*. Para os dados servirem de entrada para a rede neural, componente da aprendizagem de máquina, tais dados precisam ser válidos e estar dentro de um limite razoável para o domínio do problema, não admitindo grandes distorções.

Para ser realizado pré-processamento, primeiramente retira-se os valores não lidos pelos sensores. Esporadicamente pode ocorrer erro na leitura das informações e esse erro é salvo no arquivo de saída do sensor como um “NaN” que é a sigla para “não é

um número”. É necessário também remover os “pontos fora da curva”, conhecidos na literatura pela denominação *outliers*. É considerado *outlier* um valor muito acima ou muito abaixo dos valores padrões que estão sendo medidos. Isso pode acontecer por erros na leitura, ou algum fator externo que pode causar alteração nas leituras. Esse momento é importante pois um pequeno erro inicial que se propague, e se não seja tratado de forma correta, pode comprometer a acurácia na previsão dos dados.

Para a identificação dos *outliers* é utilizado o intervalo interquartil [Fonseca 2011] que é conhecido como método *BoxPlot*. Primeiro é multiplicada a diferença entre $Q1$ e $Q3$ por $1,5$. Sendo $Q1$ o primeiro quartil, o dado que é localizado abaixo de 25% das observações. Sendo $Q3$ o terceiro quartil, que é o dado localizado acima de 75% das observações. Após isto, são calculadas às barreiras externas. Isto é feito da mesma forma, com exceção de que a diferença entre $Q1$ e $Q3$ é multiplicada por 3 , em vez de $1,5$.

Após a identificação dos valores “NAN” e dos *outliers*, é preciso substituir esses valores por dados numericamente válidos. Para isso, é calculada a média dos valores laterais, que são o valor anterior ao *outlier* e o valor posterior, na lista de valores lidos.

A biblioteca utilizada para a implementação do módulo previsor contém, dentre outras, a função “*mlp regressor*” que é um *Regressor Perceptron Multicamada*. O *perceptron* multicamada é uma rede neural que consiste em três ou mais camadas, sendo uma camada de entrada e uma camada de saída com um ou mais camadas ocultas [das Neves 2018]. Uma vez que um MLP (Perceptron Multicamadas) é uma rede totalmente conectada. Cada nó em uma camada se conecta, com um certo peso, em um nó da camada seguinte. O módulo desenvolvido retorna um dado previsto baseado nos dados já existentes, coletados ou previstos anteriormente.

O funcionamento satisfatório da rede depende do treinamento. Tal treino é descrito como segue. É definida a matriz de entrada com duas colunas, uma de entrada e outra com o valor esperado. A primeira coluna é composta por cinco dados presentes na base de dados. Por exemplo, tendo um lista de entrada $[1,2,3,4,5,6,7]$, sendo a matriz de entrada composta por 2 linhas tal que, *linha 1* = $[1,2,3,4,5][6]$ e *linha 2* = $[2,3,4,5,6][7]$. São utilizados para treinamento todos os dados presentes na base de dados, para ter uma melhor predição. Após cada iteração do treinamento, é verificado se a taxa de sucesso atinge o valor definido padrão da função. Se sim, a rede neural fica disponível para ser usada. Se não, os pesos de entrada dos neurônios são atualizados e o processo de treinamento se repete. Existe um número máximo padrão de iterações nessa função que é de 1500 iterações. Se a taxa de sucesso não for atingida antes das 1500 iterações, o treinamento é parado.

A solução encontrada para a integração Servidor-Algoritmo de predição foi utilizar memória compartilhada. São utilizados arquivos de texto para as trocas de informações. Primeiro é usado um arquivo para os dados que o servidor envia e outro para os dados previstos e gerados pelo algoritmo de aprendizagem de máquina. No contexto de uso da aplicação/serviço, não ocorre problema de concorrência no acesso aos arquivos, pois, a ferramenta é utilizada pelo grupo específico que possui acesso à estação agrometeorológica. Contudo, um desenho de solução para controle de concorrência já existe para uso em escala maior.

4. Ferramenta

O funcionamento da ferramenta é apresentado na Figura 2. As setas representam o fluxo de execução. Não há interferência no processo de coleta dos dados. O acesso aos dados coletados é feito diretamente na base de dados, sendo ignorada esse momento no diagrama.

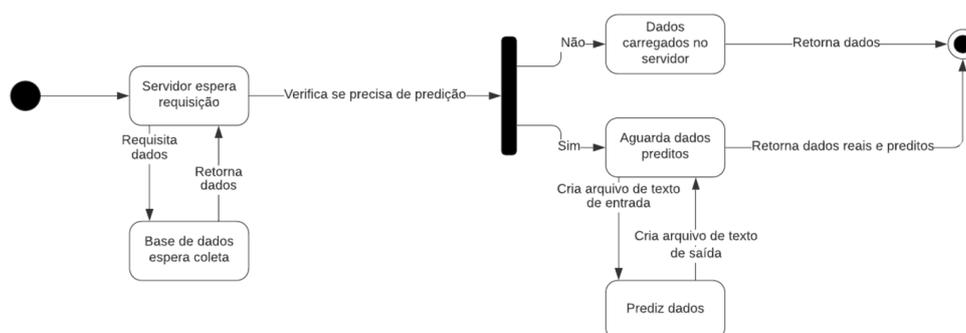


Figura 2. Diagrama de funcionamento da predição.

Fonte: Elaborada pelo autor

Começando o processo de execução do lado cliente, o usuário faz uma requisição HTTP para o servidor. Na requisição estão presentes as datas de início e fim do intervalo de tempo escolhido, e uma variável booleana indicando se vai ser preciso utilizar do módulo de predição ou não.

Ao receber a requisição, o servidor seleciona na base de dados e retorna os dados relacionados ao intervalo de tempo determinado usando uma consulta SQL. Então, é feita a verificação se é necessário utilizar o módulo predictor de dados. Se não for necessário, o servidor retorna os dados retirados da base de dados ao aplicativo móvel (cliente).

Se for necessário, o servidor faz uma verificação da quantidade de dados previstos que são necessários dado o intervalo de tempo. Em seguida, a aplicação no servidor escreve os dados em um arquivo de texto. O servidor aguarda por esses resultados, e com isso, o próprio executa o módulo predictor. A partir daí, o módulo processa os dados presentes no arquivo de texto e prevê a quantidade de dados pré-definidos, após escrever em um outro arquivo de texto todos os dados que foram previstos.

Por fim, o servidor processa o arquivo com as respostas, reorganiza os dados, e os retorna ao aplicativo. E, assim, torna-se possível a geração dos gráficos com os dados já coletados e com dados “futuros” enquanto eles ainda não são coletados. Na Figura 3 são apresentadas as telas do aplicativo, onde o usuário opta ou não pela predição.

5. Resultados

No geral os resultados obtidos nos testes foram satisfatórios para o domínio do problema. Foi identificado erro quadrático médio de 0.008 utilizando a variável *temperatura do ar* e com uma população de teste de 20 dados, definindo o erro quadrático médio como sendo a média da diferença entre o valor do estimador e do parâmetro ao quadrado [Sartóris 2013]. Há uma ressalva, a taxa de erro na predição aumenta conforme mais dados são preditos. Entretanto, a funcionalidade buscada para a ferramenta foi atingida.

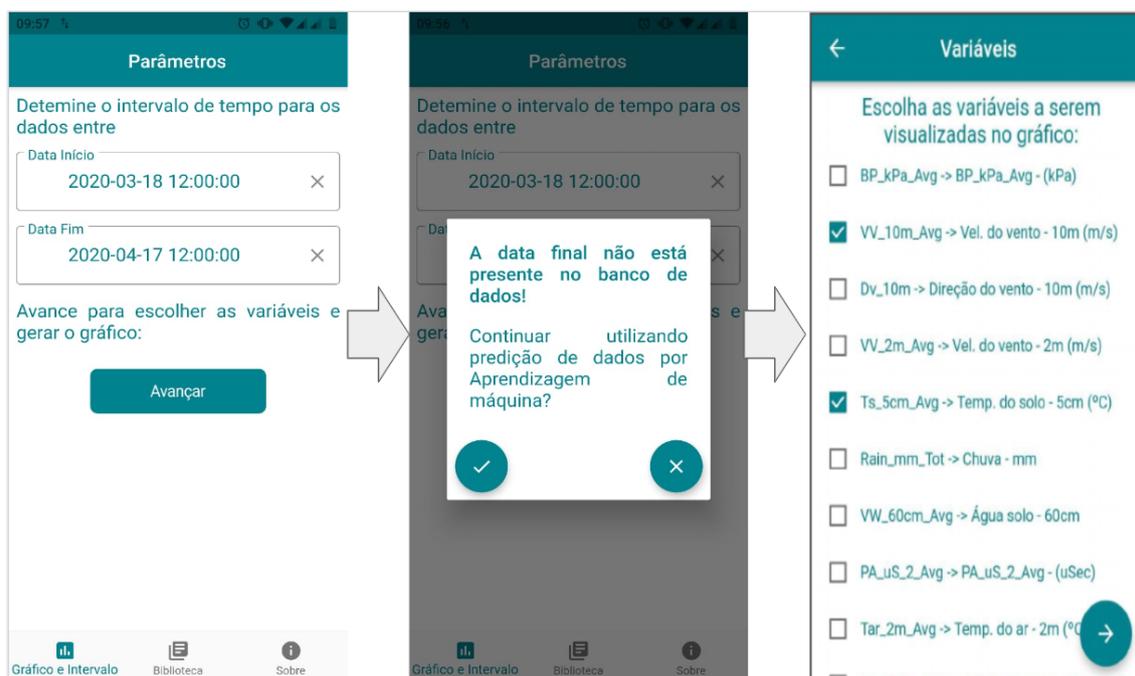


Figura 3. Processo de escolha da predição no aplicativo.

Fonte: Elaborada pelo autor

No momento em que a requisição é feita pelo cliente, o servidor é responsável por todos os processos seguintes. O próprio servidor coleta os dados na base de dados que serão utilizados e os encaminha para o módulo de predição. Após realizado o processo de predição, o servidor encaminha para o cliente os dados reais e os previstos assim sendo possível a criação dos gráficos. Os gráficos apresentam como indicação na legenda a data a partir da qual se apresentam dados previstos. As primeiras previsões têm resultados muito próximos da realidade. Porém, quanto mais longe (no futuro), os dados passam a ser previstos baseados em outros dados também previstos. Assim, essa precisão decai ao longo do tempo e o desvio aumenta.

No gráfico da Figura 4, é apresentada a diferença em relação aos dados reais coletados na estação e a predição feita pelo módulo. A linha azul representa os dados reais coletados pela estação e a linha laranja representa os dados preditos pela ferramenta para o mesmo período. O eixo x apresenta o tempo em minutos e no eixo y apresenta a média da temperatura do ar em graus celsius.

6. Conclusões

Este artigo apresentou a ferramenta de apoio à análise de dados agrometeorológicos para tomada de decisões que tende a ser promissor com seus resultados. É pretendido atender não só a comunidade acadêmica da UFAL, mas também os produtores rurais do agreste alagoano. Os dados da estação Campbell presentes na universidade são armazenados em arquivos de planilha tornando-se obsoletos com frequência. Fez-se necessária a implementação do módulo predictor de dados para viabilizar a construção de gráficos até que os dados reais sejam coletados.

Com a ferramenta funcionando, algumas limitações foram encontradas. A pre-

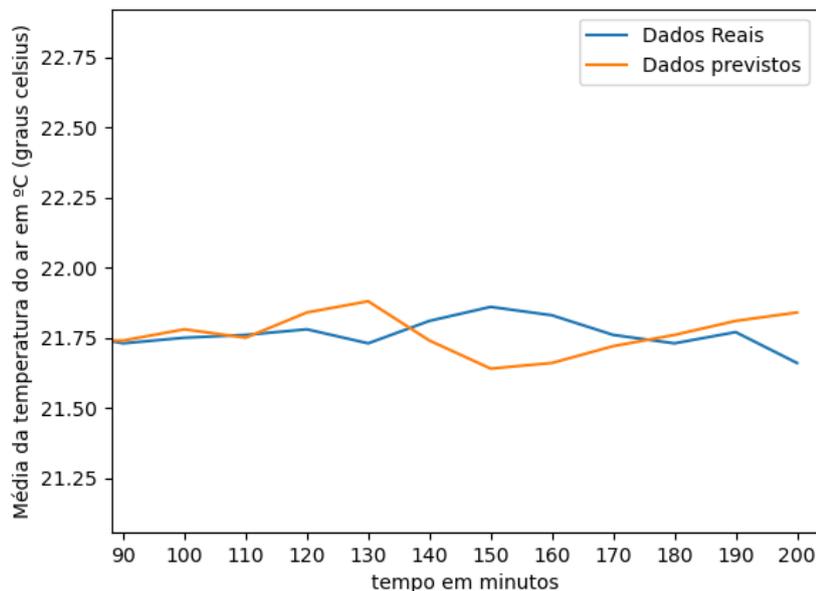


Figura 4. Comparação entre dados reais e dados previstos.

Fonte: Elaborada pelo autor

cisão na previsão dos dados decai conforme aumenta a janela de tempo escolhida pelo usuário em que não há dados armazenados na base. Outro problema é a eficiência no processamento do lado do servidor. Como os dados são armazenados e/ou previstos para intervalos de dez minutos, se o usuário definir um dia no futuro, serão previstos 144 dados. Isto representa um tempo de resposta relativamente grande.

Como trabalho futuro, é pretendido desenvolver um módulo de configuração para a leitura e carregamento do arquivo *.dat* na base de dados, desde que seja um arquivo de texto. Tal módulo servirá para mapear as variáveis climatológicas de acordo com sua sequência no arquivo. Desta maneira, outras estações disponíveis no mercado poderão fazer uso da ferramenta. Por fim, alguns ajustes e alterações simples serão feitos na interface para que a mesma se apresente mais clara e amigavelmente.

Referências

- das Neves, S. A. (2018). Técnicas de aprendizado de máquina aplicadas a classificação da qualidade de pavimentos asfálticos utilizando smartphones. Monografia (Bacharel em Engenharia da Computação), UFOP (Universidade Federal de Ouro Preto), Minas Gerais, Brazil.
- dos Santos Soares, M. (2009). Metodologias Ágeis extreme programming e scrum para o desenvolvimento de software.
- Dousseau, P. (2009). Estação agrometeorológica. <https://sigmasensors.com.br/estacao-agrometeorologica>. Último acesso em 02/03/2020.
- DSA, E. (2018). 17 casos de uso de machine learning. <https://datascienceacademy.com.br/blog/>

- 17-casos-de-uso-de-machine-learning/. Último acesso em 02/03/2020.
- FAPESP, A. (2014). Ferramenta permite análise integrada de dados sobre clima, ambiente e saúde. <http://agencia.fapesp.br/ferramenta-permite-analise-integrada-de-dados-sobre-clima-ambiente-e-19116/>. Último acesso em 02/03/2020.
- Flutter (2017). Beautiful native apps in record time. <https://flutter.dev/>. Último acesso em 02/03/2020.
- Fonseca, V. M. (2011). Detecção de outliers em dados amostrais de uma pesquisa econômica.
- Matheus, S. (2011). Ferramenta para visualização de dados meteorológicos. page 50.
- Salvador, M. (2017). Nota de pesquisa: Climap – aplicativo para análise de dados climáticos-versão 3.0. *Revista Brasileira de Climatologia*, 20:349.
- Sartóris, A. (2013). *Estatística e Introdução à Econometria*. Saraiva, São Paulo, second edition.
- Scher, S. (2018). Toward data-driven weather and climate forecasting: Approximating a simple general circulation model with deep learning. *Geophysical Research Letters*, 45(12):616–12.
- Scikit (2019). scikit-learn machine learning in python. <https://scikit-learn.org/stable/>. Último acesso em 02/03/2020.
- Seplag/AL (2015). <http://dados.al.gov.br/dataset/nucleo-de-estudos-e-projetos-daseplag/resource/bc6e26ad-d1ad-410e-baf7-9da145bffa55..> Último acesso em 02/03/2020.
- Seplag/AL (2017). Semiárido alagoano - alagoas em dados e informações. <http://dados.al.gov.br/dataset/semiarido-alagoano>. Último acesso em 02/03/2020.
- Silva, A. J. S., Júnior, A. S. d. A., and Marin, F. R. (2008). Um sistema web para a consulta de dados meteorológicos como ferramenta de apoio ao manejo de irrigação no estado do piauí. *Rev. Tecnol. Fortaleza*. <https://www.embrapa.br/informatica-agropecuaria/busca-de-publicacoes/-/publicacao/8639/um-sistema-web-para-a-consulta-de-dados-meteorologicos-como-ferrament> Último acesso em 02/03/2020.
- Spring (2020). Spring framework. <https://spring.io/projects/spring-framework>. Último acesso em 02/03/2020.
- thymeleaf (2018). Thymeleaf. <https://www.thymeleaf.org/>. Último acesso em 03/08/2020.
- Veras, E. C. e. A. M. V. B. (2011). *Capital x trabalho no campo: questao agraria, agricultura familiar e trabalho no setor sucroenergetico*. EdUFAL.
- Y. Radhika, M. S. (2009). Atmospheric temperature prediction using support vector machines. *International Journal of Computer Theory and Engineering*, 1(1).