

Explorando aprendizado de máquina com multivariáveis para previsão de enchentes em ambientes *IoT*s: um estudo empírico no sistema de monitoramento de rios E-noé

Lucas Augusto Vieira Brito¹, Danielle Bressiani, Jó Ueyama¹

¹Instituto de Ciências Matemáticas e de Computação – Universidade de São Paulo (ICMC-USP)
CEP: 13566-590 – São Carlos – SP – Brazil

lucasaugusto.vb@usp.br, danielle.bressiani@climatempo.com.br, joueyama@icmc.usp.br

Abstract. *In the last years, WSNs are being implemented in different types of applications in the urban scenario, being one of the approaches, the monitoring of natural disasters, for example floods. Typically, applications to detect natural disasters are installed in inhospitable locations and rely on multihop communication for the data to reach a sink. In this scenario, one of the main challenges of these systems is to issue warnings in a timely manner to avoid major disasters. The issue of risk and disaster management and mitigation has been discussed by several countries at the conventions on climate change and sustainable development. However, correlating WSN disseminated data to achieve this goal is not a trivial task. For these reasons, this paper proposes a prediction model based on data mining and machine learning, in order to correlate different databases to achieve a higher quality of information. The algorithm Random Forest stood out among others, reaching 97% accuracy, and outperformed them in the rigorous t-test test of the weka tool. In addition, the forecast model has shown that it can better manage the resources of an RSSF.*

Resumo. *Nos últimos anos, RSSF estão sendo implantadas em diferentes tipos de aplicações no cenário urbano, sendo um dos enfoques, o monitoramento de desastres naturais, por exemplo enchentes. Tipicamente, aplicações para detectar desastres naturais são instaladas em locais inóspitos e dependem da comunicação multihop para que os dados cheguem a um sorvedouro sink. Nesse cenário, um dos principais desafios desses sistemas é emitir alertas em tempo hábil para que evitem desastres maiores. A temática sobre gestão e diminuição de riscos e desastres tem sido discutida por diversos países nas convenções sobre mudanças climáticas e sobre desenvolvimento sustentável. Entretanto, correlacionar os dados disseminados da RSSF para alcançar tal objetivo não é uma tarefa trivial. Por esses motivos, este artigo propõem um modelo de previsão baseado de mineração de dados e aprendizado de máquina, com o intuito de correlacionar diferentes bases de dados para alcançar uma maior qualidade de informação. O algoritmo Random Forest se destacou entre os outros, atingindo 97% de acurácia, e por supera-los no teste rigoroso de avaliação test t-pareado da ferramenta o weka. Além disso o modelo de previsão, demonstrou que consegue gerenciar melhor os recursos de uma RSSF.*

1. Introdução

Os desastres naturais são causados por fenômenos de origem hidro-meteorológica, climatológica, geofísica ou biológica são problemas globais que degradam o ambiente natural, provocando danos materiais e de vítimas a um nível que excede a capacidade de auto-recuperação da comunidade local, exigindo recursos da assistência externa [Guha-Sapir et al. 2012].

No Brasil, os problemas de enchentes vêm prejudicando o país, ocasionando a perda de vidas e de bens materiais. Sendo assim, a utilização de Rede de Sensores Sem Fio (RSSF) para monitorar os rios é de grande importância, pois a poluição é uma das grandes causas de enchentes trazendo problemas ambientais através da influência na fauna e na flora presente nos rios. Com isso a RSSF pode ser benéfica para a população e para os órgãos responsáveis (Defesa Civil) para alertar sobre enchentes e sobre a poluição fluvial [PECHOTO et al. 2012].

A Rede de Sensores Sem Fio (RSSF) é composta por um nó *sink* e um conjunto de nós sensores sem fio autônomos, com recursos energéticos limitados que podem ser móveis ou fixos. São dispostos aleatoriamente em um ambiente em mudança dinâmica conhecido como campo de detecção, em que cada nó é um dispositivo de baixo consumo de energia, tendo uma comunicação sem fio e capacidade de sensoriamento de diferentes tipos de dados a partir da necessidade do ambiente [Shelke et al. 2013].

No contexto de gerenciamento e monitoramento de desastres naturais em meios urbanos, tomar uma decisão rápida e assertiva é primordial, tendo em vista os riscos à vida e ao meio ambiente [Yin et al. 2012]. Sensores físicos e outros sistemas de *IoT* (*Internet of Things*) podem realizar o monitoramento de vários pontos em uma cidade, contudo, na maioria das vezes suas aplicações são limitadas e não atendem todas regiões.

Dessa forma, RSSF é uma tecnologia vantajosa para realizar o monitoramento de ambientes naturais, já que apresentam características como: (i) simples desenvolvimento e implantação, (ii) acessibilidade em ambientes inóspitos, (iii) baixo custo, principalmente em relação à infraestrutura, (iv) possibilidade de utilização de sensores de alta precisão, (v) adaptabilidade a mudanças, (vi) escalabilidade em ambientais como afirmado por [Seal et al. 2012]. Já [Akyildiz et al. 2002] descreve a importância de se obter um melhor gerenciamento dos recursos da rede, com o objetivo de estender ao máximo a sua durabilidade, sem sacrificar a confiabilidade e a robustez do sistema. Como por exemplo a bateria, utilizando-a de forma consciente pode-se alcançar tal objetivo.

Na cidade de São Carlos- SP, Brasil, encontra-se instalada uma rede de sensores sem fio chamada E-nóe para detectar enchentes através da medição do nível do rio, desenvolvida pelo Instituto de Ciências Matemáticas e de Computação (ICMC) - Universidade de São Paulo (USP) e descrita no trabalho [PECHOTO et al. 2012]. A utilização de apenas uma única base de dados, para a detecção de enchente, limita a qualidade de informação de previsão do sistema como dito no trabalho [Furquim 2017].

Desta forma considerando a restrição da qualidade de informação que se pode obter utilizando só uma base para previsão de enchentes em uma RSSF, a proposta do trabalho é utilizar mineração de dados em varias bases. Correlacionando bases de dados do ClimaTempo, empresa que oferece serviços meteorológico, e dados do projeto E-noé tornando o sistema de previsão mais confiável. O foco é obter um maior tempo de previsão

para que medidas preventivas sejam tomadas e assim podendo evitar as consequências que os desastres de uma enchente acabam acarretando. Uma outra vantagem do modelo, foi o melhor gerenciamento dos recursos da rede, utilizando-o conforme a necessidade do ambiente.

O artigo é organizado em 4 seções: Na Seção 2 apresenta os trabalhos relacionados, a Seção 3 discute a proposta do trabalho, a Seção 4 apresenta resultados e discussão, a Seção 4.2 apresenta um estudo de caso relacionado ao modelo de predição proposto e a Seção 5 apresenta a conclusão e potenciais trabalhos futuros.

2. Trabalhos Relacionados

Esta seção apresenta trabalhos que abordam o estado da arte de como são feitos os modelos de previsão de enchentes. A revisão de literatura foi realizada com base nas pesquisas de [Seal et al. 2012, Horita et al. 2015, Wu et al. 2014, Furquim et al. 2016], sendo assim:

O Trabalho [Seal et al. 2012] descreve um esquema de previsão de inundações usando uma abordagem híbrida (centralizada e distribuída) para RSSF em rios. A arquitetura WSN consiste em vários sensores para coletar dados combinados com nós do processador, onde o algoritmo de previsão é implementado. Possui também centros de monitoramento operados manualmente, que implementam a redundância comparando a situação real com a previsão e iniciam procedimentos de evacuação.

No caso do trabalho [Horita et al. 2015], é feito um modelo de previsão que visa efetuar fusão de dados entre VGI (*Volunteered Geographic Information*) e sensores de RSSF com o objetivo de fornecer informações suplementares ao modelo de previsão para apoiar o gerenciamento de risco de inundação.

Já na pesquisa [Wu et al. 2014], é apresentada uma arquitetura para monitoramento e previsão de desastres. Esta arquitetura é baseada em RSSF e comunicação móvel que transmitem avisos e indicam as áreas em risco de desastres.

Em [Furquim et al. 2016], é apresentado um modelo de previsão que utiliza aprendizado de máquina e teoria do caos, no sentido de melhorar a previsão para evitar os danos que são causados pelas enchentes. Este modelo utiliza dados provenientes de RSSF, que representam o nível do rio e suas variações.

Portanto, com base nos trabalhos citados anteriormente, este trabalho demonstra que aplicando mineração de dados em várias bases, levando em consideração um conceito de multivariáveis, o modelo alcançou uma acurácia interessante. Alcançando assim, uma maior qualidade de informação que poderia evitar os danos que são causados pelas enchentes e um melhor gerenciamento dos recursos da rede.

3. Proposta do Trabalho

A proposta do artigo está baseada em um estudo empírico de uma RSSF chamada Enoé [PECHOTO et al. 2012] que é um projeto brasileiro para monitoramento de rios urbanos. Um dos propósitos desse projeto é poder prover condições de monitoramento pra alertas de enchentes, evitando danos maiores como perdas de vida. O seguinte artigo tem como proposta um modelo de previsão que utiliza várias fontes de dados e um sistema que utiliza multivariáveis levando em consideração fatores que interferem em enchentes,

caos e aprendizado de máquina. Porém, é relatado que umas das limitações apresentadas no trabalho foi a utilização de uma única fonte de dados, os do nível do rio, para prever inundações. Esse fato acaba limitando a acurácia e a qualidade de informação do modelo. Segundo o mesmo, a utilização de dados do nível do rio juntamente com dados de previsão do tempo aumentariam a qualidade do modelo e sua confiabilidade. Dessa forma, a proposta do artigo tem como foco o desenvolvimento de um modelo de previsão de enchentes que utiliza diversas bases de dados para obter maior confiabilidade e que produza resultados em tempo hábil, possibilitando assim que todas as medidas preventivas sejam tomadas pela Defesa Civil e evitando a possibilidade de futuros problemas nas cidades causados por inundações.

3.3. Entendimento dos dados

Essa etapa é responsável pela compreensão de dados e a matéria-prima para que esta solução seja construída. Dessa forma, o objetivo é conhecer características e limitações das bases de dados, seu histórico, sua composição e seu tipo, além de avaliar se os dados realmente são suficientes para entender o problema proposto. Um ponto relevante dessa parte é a correlação entre os dados propriamente ditos. Sendo assim, são escolhidos dados históricos que estão no mesmo período com a finalidade de encontrar anomalias e relacioná-las às condições de inundações. Os dados históricos do projeto E-noé indicam o nível a que o rio chegou e quando e onde ocorreram enchentes, enquanto os dados do ClimaTempo indicam em que momento houve chuva, visto que a chuva é a principal causa das enchentes.

3.4. Preparação dos dados

Nessa etapa, são envolvidas todas as atividades associadas à construção do conjunto final de dados, aquele que será usado na ferramenta de modelação, sofrendo inevitavelmente várias otimizações. Os dados foram modificados conforme diz [Han et al. 2011], seguindo os seguintes passos:

- **Limpeza:** eliminação de caracteres espúrios, padronização de formatos, redução de inconsistência e imputação de dados faltantes. Nesse passo, os dados foram modelados visando manter o padrão dos dados do nível do rio, para não remover as características de inundação representada nesses dados.
- **Integração:** Compilação das informações a partir de bases de dados heterogêneas.
- **Transformação:** Conversão de tipo, suavização ou remoção de ruídos, agrupamento de valores por datas, normalização de variáveis, criação de novos atributos e discretização.

3.5. Modelagem

Neste ponto serão utilizadas técnicas que sejam mais aderentes ao objetivo do projeto, seja ele uma predição, classificação, agrupamento ou regressão. A etapa de modelagem pode manter um canal de comunicação contínuo com a etapa de preparação de dados, seja para readequação dos mesmos ou para a criação de novas variáveis que ajudem a explicar o fenômeno. Nesse caso utilizamos técnicas de classificação como por exemplo SVM (*Support Vector Machine*), *Random Forest*, *Multilayer Perceptron* e J48 (Algoritmo de árvore de decisão).

3.6. Avaliação

Considerada uma das etapas mais críticas do modelo, pois tem como finalidade avaliar a utilidade do modelo construído. Diversas ferramentas gráficas são utilizadas para visualização e análise de resultados. Sendo uma forma de avaliar o quanto de confiabilidade o modelo consegue atingir, alguns exemplos são:

- *Cross Validation*;
- *Supplied test set*;
- *Use training set*;
- *Percentage split*;

É primordial rever os passos executados em sua construção e verificar se os mesmos permitem atingir os objetivos propostos na etapa de análise do problema.

3.7. Aplicação do Modelo

Nesta etapa, o projeto é encerrado com a entrega. É a etapa menos técnica do processo de mineração de dados, mas não a menos importante.

4. Resultados e Discussões

Para o desenvolvimento do modelo de previsão de enchentes, proposto pelo artigo, algumas características serão ressaltadas. Foram utilizados os dados do nível do rio (E-noé) como parâmetro para determinar quando inundações devem ou não ocorrer. A partir disso, serão utilizados algoritmos de classificação para a confecção do modelo. No projeto E-noé existem vários pontos de sensoriamento espalhados pelo rio, assim foi decidido analisar o ponto específico instalado no *shopping* de São Carlos-SP. A escolha foi feita pois esse ponto é considerado o local em que ocorrem maior número de inundações, devido à sua localização - a união de dois córregos. Segundo [Furquim 2017], especialistas definem que nesse local, caso o nível do rio ultrapasse os 200 centímetros (2 metros) já é considerado um caso de enchente.

Em relação a modelagem dos dados, na base do ClimaTempo foram escolhidas algumas variáveis que segundo especialistas da área de meteorologia indicam chuva, são elas: (i) Temperatura máxima, (ii) Temperatura Mínima, (iii) Umidade, (iv) precipitação e (v) Intensidade do vento. Já na base do projeto E-nóe, foi utilizado a variável que mede o nível do rio. Após essa fase, foi feita a correlação das duas bases (Projeto E-nóe e Climatempo), criando-se uma nova base com todas as variáveis modeladas em uma mesma janela e com um mesmo período de tempo. A janela que foi escolhida, levou em consideração a busca de se ter o equilíbrio entre as classes de **casos de enchentes e não enchentes**. Pois, foi observado nos históricos dos dados do E-nóe, a característica de se ter mais casos de uma única classe (não enchente) o que poderia interferir no aprendizado do modelo.

Outro ponto importante, foi a decisão de analisar o mês de novembro de 2015, pois foi o mês onde o rio atingiu seu maior ápice de inundações chegando a mais de 600 centímetros (6 metros) de altura. A **Figura 2** demonstra em detalhes como a altura do nível do rio varia dia a dia. Vale ressaltar que a média do mês de novembro chegou a 93 centímetros, uma média considerada alta em comparação com os outros meses. A partir disso, foi decidido analisar isoladamente as datas do mês de novembro em que a altura do

Tabela 1. Casos de Enchentes

Data	Altura do Rio
15/11/15	Acima de 400 cm (4 metros)
20/11/15	Acima de 200 cm (2 metros)
23/11/15	Acima de 600 cm (6 metros)

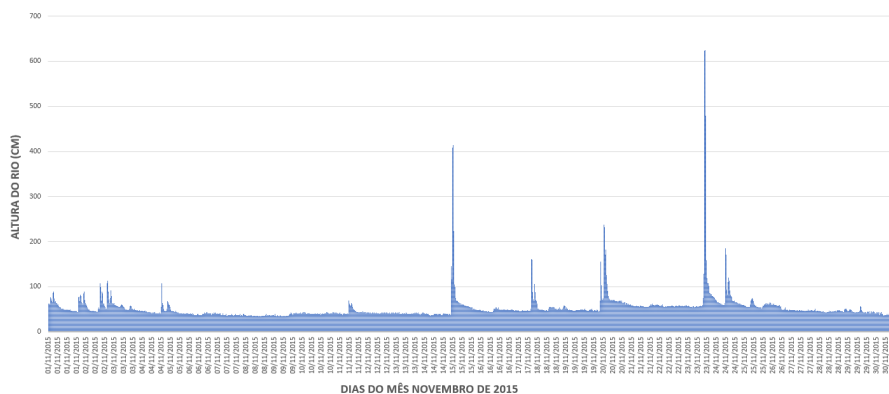


Figura 2. Gráfico da variação do nível do rio do mês de novembro de 2015

rio ultrapassou os 200 centímetros (2 metros). A **Tabela 1** foi elaborada para apresentar essa lista de datas em que a altura do rio já qualificava o caso como enchente.

Assim, levando em consideração o caso mais grave dentre 3 datas listadas, o rio alcançou seu maior nível no dia 23/11/15, motivo pelo qual foi decidido analisar essa data separadamente, para maior entendimento do padrão alcançado, e tornando-a a escolhida como base de treino para análise dos algoritmos.

A **Figura 3** apresenta a voracidade que o rio sobe conforme analisamos do ponto **1** em que o rio está a 94,2 centímetro de altura ao ponto **2** que chega a 624,86 centímetro de altura, representando assim um aumento de 665% em um curto espaço de tempo (45 minutos aproximadamente), tornando essa anomalia interessante para o estudo.

4.1. Modelo de previsão: Análise dos Algoritmos de Aprendizado de Máquina

Para a avaliação da escolha do melhor algoritmo utilizado no modelo proposto, foram considerados conceitos como performance e flexibilidade. Sendo assim, foi utilizada a ferramenta Weka [Hall et al. 2009] para análise desses algoritmos. Dentre todos os algoritmos disponíveis na ferramenta, foram avaliados 4 no âmbito de classificação. Os algoritmos escolhidos e as características dos dados que foram utilizadas para elaboração do modelo de previsão são:

- Algoritmo baseado em função:
 - *Multilayer Perceptron*;
 - *Support Vector Machine (SMO)*;
- Algoritmo baseado em árvore de decisão:
 - J48 (Árvore de Decisão);
 - *Random Forest*;

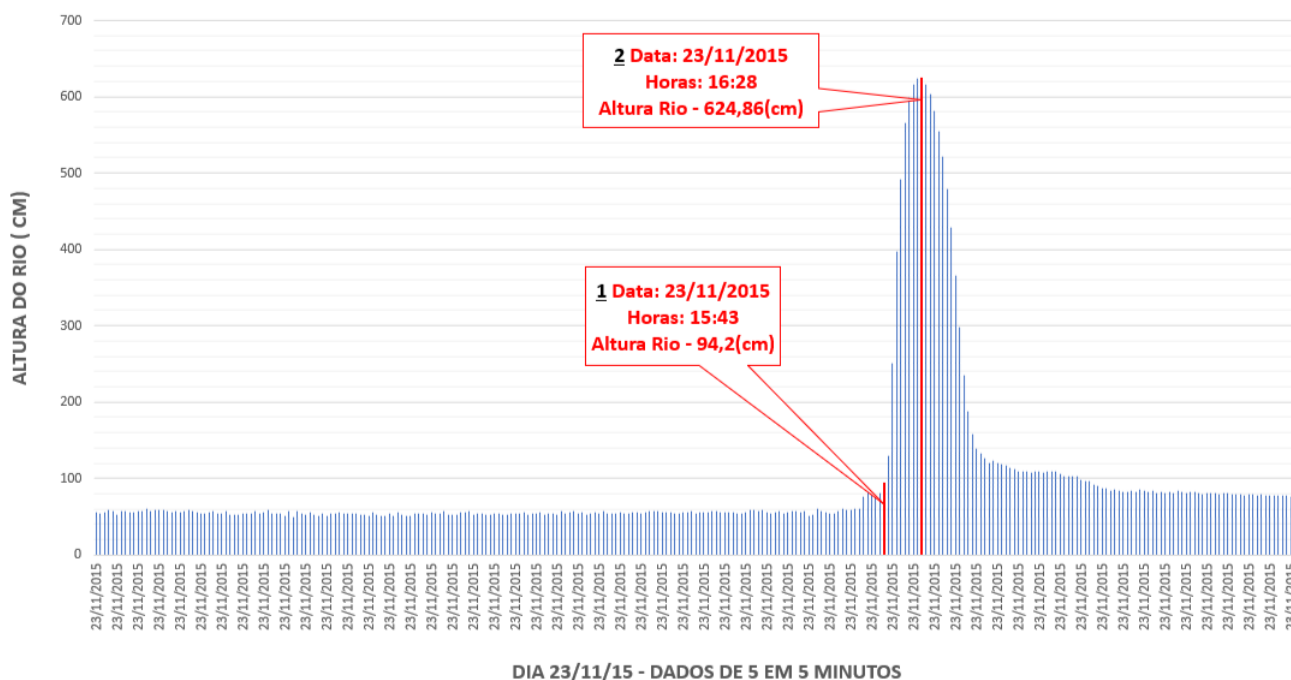


Figura 3. Gráfico de análise de enchente do dia 23/11/15

- Configuração do algoritmos: Todos os algoritmos não sofreram modificações, assim seguiram o padrão *default* no *Weka*;
- **Bases de Dados:** Foram utilizados dados do projeto E-noé (Nível do rio) e dados do ClimaTempo como: I) Temperatura; II) Umidade Relativa; III) Precipitação; IV) Pluviômetro e V) Intensidade de vento;

As bases de dados históricas, são importante para treinar o modelo. Dessa forma, sofreram uma modelagem e foram analisadas em um mesmo período de tempo pré-definido. O balanceamento das variáveis foi essencial para se obter quantidades idênticas de **casos de enchentes** e **não enchentes** não influenciando a precisão do modelo proposto.

Para a avaliação do desempenho dos algoritmos, foi escolhido o método de *Cross Validation 24 folds*, sendo que o mesmo divide as instâncias em 24 partes, obtendo-se 24 combinações de treinamento-teste (23 partes para treinamento e 1 para teste). No caso do nosso modelo, essa forma de treinamento e teste evitam que ocorra *overfitting* e que se aumente a generalização do modelo. Levando em consideração que no dia 23/11/15, temos a maior enchente que ocorreu no ano de 2015, definimos assim utilizar dados do projeto e-noé(nível rio) e dados do ClimaTempo desse dia, com a estratégia de alcançar o objetivo de conseguir um modelo de previsão com maior qualidade de informação.

No sentido de classificação dos algoritmos, foi utilizado o **teste t-pareado**, que é um teste padrão no *Weka*. Uma das características desse teste é considerar as dependências entre as amostras sendo um fator importante para avaliação do modelo. O intervalo definido para ele foi de 95% levando em consideração o padrão do *Weka*.

Para efetuar as avaliações, o **teste t-pareado** utiliza as análises de duas variáveis

que são:

- **Análise ROC (Receiver Operating Characteristic):** Consiste em um método gráfico para avaliação, organização e seleção de sistema de diagnósticos ou predição. Recentemente essa métrica foi adicionada em aprendizagem de máquina e mineração de dados como uma ferramenta útil e poderosa para a avaliação de modelos de classificação.
- **Matriz de confusão:** É uma matriz de valores reais e valores preditivos pelo seu classificador. Ela permite verificar se seu modelo de classificação está mantendo a generalização e conseguindo classificar todo os dados, não só uma única classe.

Tabela 2. Análise Matriz de Confusão

Acurácia	Erro	Classificadores
97.619	2.381	Support Vector Machine(SMO)
97.619	2.381	Random Forest
92.857	7.143	Multilayer Perceptron
97.619	2.381	J48 (Árvore de Decisão)

Analisando a **Tabela 2**, podemos observar que os 3 algoritmos (*Support Vector Machine*, *Random Forest* e J48) tiveram a mesma porcentagem de acurácia na questão de classificação do modelo, com base nos dados modelados do dia 23/11/15 com *Cross Validation 24 folds*. Ressaltando que as configurações dos algoritmos foram deixadas no modo *default* da ferramenta Weka, assim a matriz de confusão dos três algoritmos obtiveram a mesma taxa de acerto em ambas as classes. Dessa forma, a análise *ROC* que vai decidir qual algoritmo é mais eficiente para o modelo.

Tabela 3. Análise Variável ROC

Classificadores	ROC (Receiver Operating Characteristic)
<i>Support Vector Machine (SMO)</i>	0,976
Random Forest	0,998
<i>Multilayer Perceptron</i>	0,976
J48 (Árvore de Decisão)	0,995

Já na **Tabela 3**, é possível ver a diferença entre *ROC* de cada algoritmo. É definido pela análise que o valor que estiver mais próximo de 1 é o algoritmo com maior taxa de TP (*True Positive*) e menor taxa de FP (*False Positive*). Consequentemente, o algoritmo que se destacou na análise foi o **Random Forest**, que por uma pequena diferença em relação os outros, é considerado o mais apto para o modelo de previsão.

4.2. Estudo de Caso: Gerenciamento dos Recursos

Nessa seção foram estudados os benefícios de um modelo de previsão de enchente. Segundo [Akyildiz et al. 2002] energia, memória, recursos de computação e alcance de transmissão são as limitações da RSSF. Sendo assim, com o nosso modelo de predição de enchentes ajustado que utiliza bases de dados do ClimaTempo e do E-noé, podemos não só avisar sobre a possibilidade de enchente, mas também definir como melhor gerenciar seus recursos, conforme necessidade do ambiente. Aplicando assim, o conceito de **adaptabilidade** da rede no uso oportuno de seus recursos. Com o objetivo de evitar futuras percas por falta de recursos da RSSF, aumentando assim o seu tempo de vida.

Para exemplificar melhor o quanto a rede poderia ser melhorada, foi definido abordar no momento o conceito de energia da rede, visto que a troca de mensagens é o fator que mais consome energia em RSSF. Sendo assim, foram realizadas 30 execuções de um mesmo cenário de simulação no programa OMNeT++, com o auxílio do pacote MiXiM. O cenário simulado baseia-se em uma RSSF composta por diversos nós de sensores trocando mensagens entre si e entre o nó sink.

Nesse contexto algumas características são importantes serem ressaltadas. Os nós de sensores simulados são baseados em sensores reais, usando baterias com capacidade de 1.000 mAh (miliampère-hora), e comunicando-se com o uso do protocolo ZigBee (IEEE 802.15.4) – protocolo que tem como objetivo principal permitir níveis de confiabilidade na comunicação sem fio, sem deteriorar o uso da bateria. A RSSF com modelo de predição tem a vantagem de prever com antecedência, quando irá ocorrer uma enchente. Assim, ela utiliza desse artifício para alterar a rede conforme sua necessidade. As duas configurações são:

- **Alto Risco:** É caracterizado pelo modelo de previsão como um ambiente hostil, pelo fato da classificação dos dados recebidos em tempo real determinar que a probabilidade de enchente é alta, então a arquitetura tem que funcionar de forma contínua sem interrupção. Assim o comportamento dessa arquitetura é baseado em funcionar da seguinte maneira, recolhe os dados do ambiente monitorado através dos nós sensores, essa informação é enviada para nuvem através do nó sink. Assim o fluxo de informações de dados nesse cenário é alto. Com essas informações é possível alertar toda população ribeirinha, defesa civil e órgãos responsáveis sobre problemas que podem vir acontecer provenientes das enchentes, salvando vidas e prevenindo assim desastre maiores.
- **Baixo Risco:** É caracterizado pelo modelo de previsão, como um modo de risco baixo, pelo fato da classificação dos dados recebidos em tempo real determinar que o ambiente está favorável. Dessa forma o comportamento dessa arquitetura é baseada em economia de energia, só o nó sink fica ativo para comandar as ações e o outros nós sensores entram em modo **Sleep** (Hibernação), ficando assim nesse estado por um determinado tempo. Após ter decorrido esse tempo, é feita uma nova verificação em tempo real desse cenário podendo mudar ou não dependendo das condições do ambiente monitorado e da classificação de probabilidade de enchente. Nesse cenário o fluxo de informações é baixo. O resultado esperado é um uso eficiente de energia e um menor gasto energético.

Sendo assim, a **Figura 4** apresenta o gasto das duas plataformas simuladas em um período de 24 horas. Enquanto a RSSF tradicional gasta cerca de 72% da carga de bateria, a RSSF com modelo de predição gasta 48% no mesmo período, economizando cerca de 24%. Essa economia foi devido à programação do modo Sleep, configurado para que os nós entrassem no estado de economia de bateria a cada 300 segundos. Esse modo simula o desligamento do rádio dos nós de sensores, sendo utilizado *somente* quando o ambiente não está em risco, ou seja, no modo de **baixo risco**, considerando a baixa probabilidade de um desastre natural ocorrer nos próximos segundos no local. Dessa forma, os nós de sensores economizam bateria para momentos em que a troca de mensagens será essencial.

Se for considerada a simulação para 48 horas com as mesmas características descritas acima, a arquitetura de RSSF com modelo de previsão funcionaria normalmente,

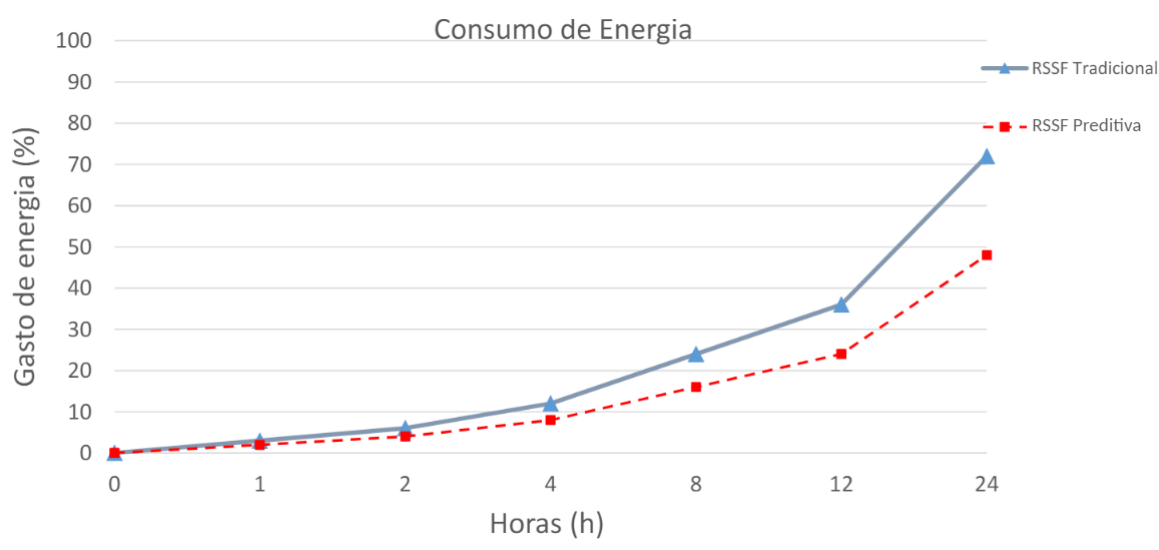


Figura 4. Comparação Consumo de Energia em uma RSSF

enquanto que a arquitetura da RSSF tradicional pararia de funcionar, pois seus nós esgotariam a bateria. Assim, a principal contribuição desse trabalho, foi o desenvolvimento de um modelo de previsão de enchentes utilizando um método de mineração de dados em várias bases de dados (ClimaTempo e Dados do E-noé), alcançando maior qualidade de informação e atingindo maior tempo de previsão. Outro benefício do modelo de previsão é a possibilidade de adaptabilidade da arquitetura de RSSF, gerenciando de maneira mais eficiente seus recursos e aumentando o tempo de vida da rede.

5. Conclusões

Este artigo tem como objetivo contribuir para o conceito de cidades inteligentes, ou seja, cidades monitoradas em todos os seus aspectos com o foco de melhorar sua infraestrutura. Tal monitoração é realizada em tempo real utilizando diversas tecnologias, tornando centros urbanos melhores para se viver. O principal objetivo deste trabalho, é o desenvolvimento de modelo de previsão que utiliza várias bases de dados como aprendizado se tornando assim mais confiável. Um estudo de caso foi apresentado, considerando o modelo de previsão desenvolvido visando um maior gerenciamento dos recursos da RSSF. O artigo tem o papel importante de fortalecer cada vez mais a ideia de um ambiente inteligente que utiliza Inteligência artificial e bases de dados para tomar decisões.

Resultados mostram que o algoritmo *Random Forest* foi o que obteve maior aceitação na avaliação dentre outros algoritmos, alcançando uma taxa de acurácia de aproximadamente 97% e passando com probidade em relação à análise *ROC* e à análise de matriz de confusão. Já no estudo de caso apresentado, a arquitetura de RSSF com o modelo preditivo desenvolvido, obteve aproximadamente 24% de economia na bateria comparada com uma RSSF tradicional sem deixar de ser eficiente na troca de mensagens entre seus nós. Para isso, os nós de sensores continuamente monitoram as características do ambiente atual, e mudam de comportamento visando a economia de bateria, sem prejudicar o bom funcionamento da rede.

Como trabalhos futuros, pretende-se adicionar mais conhecimento ao modelo de

previsão, fazendo com que reconheça todas as outras vertentes dos dados, sendo esses não apenas os de um dia de previsão, mas também permitir que compreenda dados históricos do Climatempo e do E-noé de anos anteriores, permitindo assim maior capacidade de tomar decisão, maior generalização e tornando o modelo cada vez mais confiável. Medir também o tempo de previsão em uma simulação com dados nunca vistos. Outra vertente para melhorar o trabalho, seria obter uma maior adaptabilidade no sistema, considerando todos os recursos disposto da RSSF.

Referências

- Akyildiz, I. F., Su, W., Sankarasubramaniam, Y., and Cayirci, E. (2002). A survey on sensor networks. *IEEE Communications magazine*, 40(8):102–114.
- Furquim, G., Pessin, G., Faiçal, B. S., Mendiondo, E. M., and Ueyama, J. (2016). Improving the accuracy of a flood forecasting model by means of machine learning and chaos theory. *Neural computing and applications*, 27(5):1129–1141.
- Furquim, G. A. (2017). *Uma abordagem tolerante a falhas para a previsão de desastres naturais baseada em IoT e aprendizado de máquina*. PhD thesis, Universidade de São Paulo.
- Guha-Sapir, D., Vos, F., Below, R., and Ponserre, S. (2012). Annual disaster statistical review 2011: the numbers and trends. Technical report, Centre for Research on the Epidemiology of Disasters (CRED).
- Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The weka data mining software: an update. *ACM SIGKDD explorations newsletter*, 11(1):10–18.
- Han, J., Pei, J., and Kamber, M. (2011). *Data mining: concepts and techniques*. Elsevier.
- Horita, F. E., de Albuquerque, J. P., Degrossi, L. C., Mendiondo, E. M., and Ueyama, J. (2015). Development of a spatial decision support system for flood risk management in brazil that combines volunteered geographic information with wireless sensor networks. *Computers & Geosciences*, 80:84–94.
- PECHOTO, M. M., UEYAMA, J., and Pereira, J. (2012). E-noé: Rede de sensores sem fio para monitorar rios urbanos. In *Congresso Brasileiro Sobre Desastres Naturais*.
- Reis, L. P., Vieira, J., Lemos, P., Novais, R., and Faria, B. M. (2017). Higher education access prediction using data-mining. In *Information Systems and Technologies (CISTI), 2017 12th Iberian Conference on*, pages 1–8. IEEE.
- Seal, V., Raha, A., Maity, S., Mitra, S. K., Mukherjee, A., and Naskar, M. K. (2012). A simple flood forecasting scheme using wireless sensor networks. *arXiv preprint arXiv:1203.2511*.
- Shelke, R., Kulkarni, G., Sutar, R., Bhore, P., Nilesh, D., and Belsare, S. (2013). Energy management in wireless sensor network. In *Computer Modelling and Simulation (UK-Sim), 2013 UKSim 15th International Conference on*, pages 668–671. IEEE.
- Wu, C.-I., Kung, H.-Y., Chen, C.-H., and Kuo, L.-C. (2014). An intelligent slope disaster prediction and monitoring system based on wsn and anp. *Expert Systems with Applications*, 41(10):4554–4562.

Yin, J., Lampert, A., Cameron, M., Robinson, B., and Power, R. (2012). Using social media to enhance emergency situation awareness. *IEEE Intelligent Systems*, 27(6):52–59.