DOI: 10.5748/19CONTECSI/PSE/HEA/7052

ANÁLISES ESTATÍSTICAS E COMPUTACIONAIS EM UMA REDE DE COMPUTADORES HOSPITALAR

Josué Batista Matos Deschamps de Melo ; https://orcid.org/0000-0001-8978-1664
Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

Higor Viana de Morais ; https://orcid.org/0000-0001-9785-1925
Senac Osasco

Clarice Gameiro da Fonseca Pachi ; https://orcid.org/0000-0001-7833-6215
Centro Universitário Senac Santo Amaro

Jorge Futoshi Yamamoto ; https://orcid.org/0000-0003-0360-2800
Hospital das Clínicas da Faculdade de Medicina da Universidade de São Paulo

Análises Estatísticas e Computacionais Em Uma Rede de Computadores Hospitalar

Statistical and Computational Analysis in a Hospital Computer Network

ABSTRACT

Advances in transmission technology have allowed computer networks to exchange information on a large scale at speeds compatible with the needs of the information society and data communication. In this way, issues related to infrastructure became more visible and of interest to managers from different disciplines.

Knowing the behavior of the network or its segments is important to know which protocols consume more resources. An important aspect of any network monitoring task is precisely related to choosing the best method to perform it. Computer networks can be used to transmit information such as images, audio, video, and various types of files.

This transmission can be used for a variety of purposes, including the transmission of medical images using certain technologies such as DICOM and PACS. There are protocols that allow the integration between different medical image acquisition systems. DICOM is a communication system for digital images in medicine, represented by a set of rules created by the American Radiological Society in 1983 to facilitate the storage and communication of images. PACS is a type of technology that facilitates the communication and archiving of image files.

In this context, the proposal was born in partnership with NETI - HCFMUSP, which aims to analyze the data flow of the computer network using open-source tools.

The objective of this work is to analyze and interpret the data composed in these communication systems, where the flow of information carried by the hospital's computer network will be treated and visualized with the use of mathematical and computational tools.

Keywords: network, hospital, system

RESUMO

Os avanços na tecnologia de transmissão permitiram que as redes de computadores trocassem informações em larga escala em velocidades compatíveis com as necessidades da sociedade da informação e da comunicação de dados. Dessa forma, as questões relacionadas à infraestrutura tornaram-se mais visíveis e de interesse de gestores de diferentes disciplinas.

Conhecer o comportamento da rede ou de seus segmentos é importante para saber quais protocolos consomem mais recursos. Um aspecto importante de qualquer tarefa de monitoramento de rede está justamente relacionado à escolha do melhor método para realizá-la. As redes de computadores podem ser usadas para transmitir informações como imagens, áudio, vídeo e vários tipos de arquivos.

Essa transmissão pode ser usada para diversos fins, incluindo a transmissão de imagens médicas usando certas tecnologias, como DICOM e PACS. Existem protocolos que permitem a integração entre diferentes sistemas de aquisição de imagens médicas. O DICOM é um sistema de comunicação para imagens digitais em medicina, representado por um conjunto de regras criado pela American Radiological Society em 1983 para facilitar o armazenamento e a comunicação de imagens. O PACS é um tipo de tecnologia que facilita a comunicação e o arquivamento dos arquivos de imagens.

Nesse contexto, a proposta nasceu em parceria com o NETI - HCFMUSP que tem como objetivo analisar o fluxo de dados da rede de computadores utilizando ferramentas de código aberto.

O trabalho tem como objetivo analisar e interpretar os dados compostos nestes sistemas de comunicação, onde o fluxo de informações transitadas pela rede de computadores do hospital será tratado e visualizado com o uso de ferramentas matemáticas e computacionais.

Palavras-chave: rede, hospital, sistema

INTRODUÇÃO

As redes de computadores tornaram-se parte essencial do dia a dia das pessoas e empresas. Apesar dessa importância, os processos de gerenciamento de redes ainda não foram totalmente explorados. O monitoramento do tráfego de rede, um dos muitos processos sob responsabilidade do gerenciamento de rede, não é exceção. Conhecer o comportamento da rede ou de seus segmentos é importante para saber quais protocolos consomem mais recursos. (Comer, 2007)

Um aspecto importante de qualquer tarefa de monitoramento de rede está justamente relacionado à escolha do melhor método para realizá-la. Isso porque o monitoramento pode criar condições artificiais que não retratam a realidade do tráfego, mas ainda são úteis para analisar as informações obtidas através desse método.

Dessa forma, questões relacionadas ao fluxo de dados ficaram mais claras e interessantes. As redes de computadores podem ser usadas para transmitir informações como imagens, áudio, vídeo e vários tipos de arquivos. Essa transmissão pode ser usada para diversos fins, incluindo a transmissão de imagens médicas usando certas tecnologias, como DICOM (Digital Imaging and Communications in Medicine) e PACS (Image Archiving and Communication Systems).

O objetivo da imagem médica é permitir que os profissionais da área da saúde olhem de forma não invasiva dentro do corpo humano para fazer diagnósticos precisos. A tecnologia de raios X existe desde o início do século XIX, com a invenção da tomografia computadorizada em 1972 e o advento da imagem médica tridimensional (Yoo, 2004).

Existem protocolos que permitem a integração entre diferentes sistemas de aquisição de imagens médicas, sendo o mais conhecido o HL7 (Health Level 7). Isso possibilita a troca de informações entre os sistemas de informação do paciente e o DICOM, e é o material de pesquisa deste trabalho (Pianykh, 2012).

O DICOM é um sistema de comunicação para imagens digitais em medicina, representado por um conjunto de regras criado pela American Radiological Society em 1983 para facilitar o armazenamento e a comunicação de imagens (Pianykh, 2012).

O PACS é um tipo de tecnologia que facilita a comunicação e o arquivamento dos arquivos de imagens. É uma ferramenta muito útil para evitar a perda de relatórios, reduzir custos e aumentar a flexibilidade dos dados armazenados (Strickland, 2000).

Existem soluções para ajudar nas tarefas de análise de tráfego de rede de computadores. Essas ferramentas geralmente permitem coletar e armazenar dados para análise em tempo real e posterior. Ao usar essas ferramentas em pontos estratégicos da rede, pode-se capturar informações ou fragmentos de dados, dependendo das necessidades analíticas. A escolha dos pontos de coleta de tráfego é, portanto, parte essencial de qualquer projeto de análise de rede, e a tarefa de coletar dados de tráfego torna-se cada vez mais complexa à medida que as redes crescem (Macedo, 2015).

Nesse contexto, nasceu proposta em parceria com o NETI - HCFMUSP (Núcleo Especializado em Tecnologia da Informação do Hospital das Clínicas Faculdade de Medicina da Universidade de São Paulo) que forneceu dados com o uso de ferramentas de computação relacionadas à programação, como a linguagem Python e suas bibliotecas. Usando ferramentas de desenvolvimento, é possível utilizar fórmulas estatísticas e elementos de comparação para garantir uma análise consistente dos dados.

METODOLOGIA

As análises são realizadas com o uso de vários tipos de modelos matemáticos e o auxílio da computação para tratar grandes volumes de dados. As aplicações na pesquisa de redes de computadores possuem um papel fundamental.

Os métodos utilizados nos estudos variaram, sendo realizado um levantamento bibliográfico como passo inicial. Nesse sentido, Luna (1997, p. 20) explica: "(...) o objetivo da revisão bibliográfica é constatar a situação real da área de pesquisa, conhecer o que se sabe até hoje, quais as falhas já constatadas e quais os principais obstáculos teóricos e/ou metodológicos."

Na segunda etapa, foram analisados os dados do NETI-HCFMUSP para verificar o comportamento do fluxo de dados.

Os dados disponíveis na amostra foram capturados através de softwares disponíveis pelo setor de pesquisas do Núcleo Especializado em Tecnologia de Informação (NETI), no qual estes dados pudessem ser enviados, visualizados, organizados e manipulados. Essa interação com o setor de pesquisa é crucial, pois somente assim é possível possuir uma base de dados do fluxo da rede de computadores.

Mesmo com toda a facilidade de comunicação e acesso a tais informações, existem restrições quanto aos tipos de dados que podem ser lidos, pois deve-se respeitar normas impostas pela COMISSÃO DE ÉTICA EM PESQUISA EM SERES HUMANOS DA FACULDADE DE MEDICINA DA UNIVERSIDADE DE SÃO PAULO – CEP.

Dessa forma os dados obtidos na amostra se davam pelo período de 22 de fevereiro de 2021 a 1º de março de 2021. Os dados permitem uma visualização da quantidade de informações num formato de planilhas eletrônicas.

As informações da amostra foram dispostas em colunas no Excel e cada coluna mostrava:

- Porta: revela o número da porta do Switch em que os dados são transitados pela rede;
- DIO "Distribuidor Interno Óptico": é um endereço do conector óptico no qual se encontra;
- Velocidade: mostra qual a velocidade em gigabytes por segundo na transferência de dados pela rede, ou seja, a taxa de transferência;
- Alteração: indica a data e horário em que houve tal alteração pela rede.

Figura 1: Captura de tela dos dados organizados por endereço de IP

Tipo de Protocolo	DICOM
Quantidade	7256

IP Source	Quantidade
10.65.136.125	92
10.65.137.46	1413
10.65.142.66	767
10.65.136.122	101
10.65.136.120	63
10.65.142.84	1354
10.65.142.10	11
10.65.136.123	68
10.65.136.125	88
10.65.136.124	94
10.65.142.81	391
10.65.142.83	571
10.65.142.35	170
10.65.142.103	32
10.65.142.65	401
10.65.142.87	28
10.65.136.126	111
10.65.142.88	1
10.65.142.98	10
10.65.142.64	429
10.65.142.56	20
10.65.142.165	36
10.65.142.54	36
10.65.142.166	38
10.65.136.121	94

IP Destination	Quantidade
10.65.142.36	348
10.65.142.83	73
10.65.137.46	4958
10.65.142.65	47
10.65.143.26	51
10.65.142.37	275
10.65.143.3	69
10.65.142.66	38
10.65.142.57	1
10.65.142.18	59
10.65.143.29	68
10.65.142.166	2
10.65.143.92	18
10.65.142.84	63
10.65.143.27	30
10.65.143.224	23
10.65.142.24	15
10.65.143.2	143
10.65.143.28	46
10.65.142.87	3
10.65.143.225	8
10.65.142.88	1
10.65.143.5	109
10.65.143.4	108
10.65.143.7	88

Fonte: Elaborado pelos autores (2022)

A Figura 1 mostra um resumo dos dados separados por endereço de IP, onde se encontra o IP de origem (IP Source) e o IP de destino (IP Destination), e "quantidade" representa o número de vezes em que houve um tráfego pela rede de determinado endereço de IP com o uso do DICOM.

Indiretamente, os métodos de aprendizado de máquina são usados de maneira simplificada, ou seja, uma parte de inteligência artificial, pois permite ao fazer previsões, a saída do código que segue uma sequência numa espécie de “treino”.

Existem outras técnicas de aprendizado de máquina com desempenho superior ao trabalhar fora dos modelos ARIMA.

No entanto, o modelo ARIMA ainda usa das melhores previsões podem ser compreendidas e analisadas. O erro entre os dados previstos e reais pode ser maior, dependendo dos parâmetros do modelo, tamanho da amostra e quantidade de dados que serão calculados para a previsão.

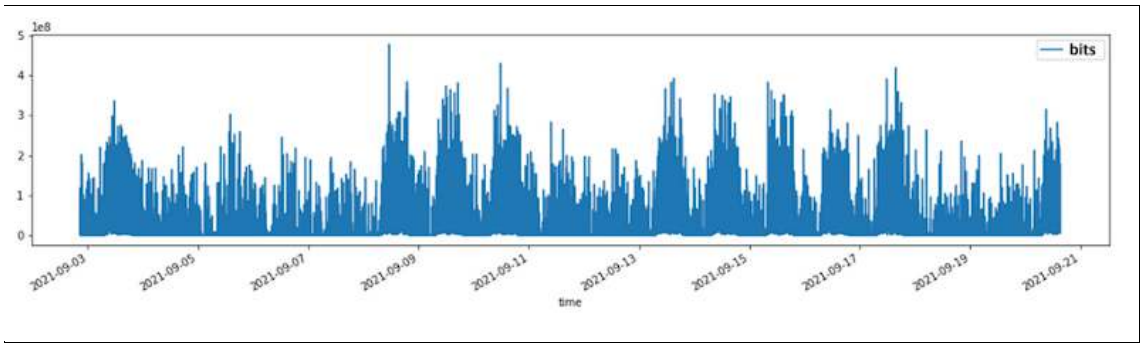
Neste trabalho o tamanho da amostra é pequeno, então o erro tende a ser grande, portanto, é necessário que encontre parâmetros adequados para tornar o erro menor.

RESULTADOS

A análise através da rede pode ser visualizada através de gráficos. No caso de uma rede de computadores deve-se analisar dois tipos de fluxos: o fluxo de entrada de dados (ato em que os dados são enviados/inseridos nos sistemas durante o transporte pela rede) e o fluxo de saída de dados (ato em que os dados são receptados/consultados nos sistemas durante o transporte pela rede).

Os dados são analisados de acordo com a transição entre o PACS01 (equipamento que armazena as imagens médicas) e CORE02 (Switch principal da rede).

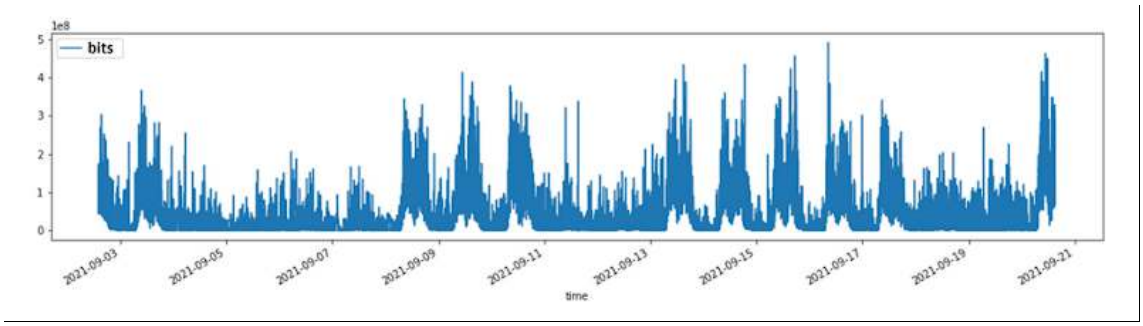
Gráfico 1: Tráfego In PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

O Gráfico 1 mostra o comportamento da rede nos momentos de inserção de dados, onde time é o tempo em questão com as respectivas datas de ocorrência e bits se trata da quantidade computacional dos dados.

Gráfico 2: Tráfego Out PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

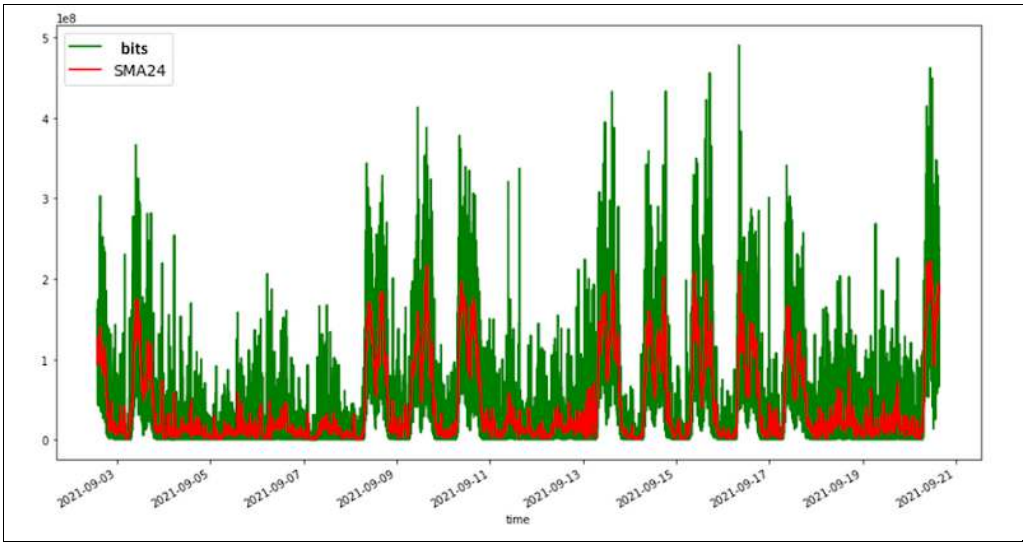
O Gráfico 2 mostra o comportamento da rede nos momentos de consulta de dados, onde time é o tempo em questão com as respectivas datas de ocorrência e bits se trata da quantidade computacional dos dados.

Durante a análise é possível notar alguns picos que se trata do horário em que o tráfego ocorre, pois em horários comerciais há tendência de maior fluxo de dados.

Com estas informações, é possível manipular os dados com o uso de ferramentas estatísticas e o auxílio da linguagem de programação.

Dessa forma, o uso de médias móveis revela não só como os dados se comportam, mas também evidenciar o crescimento e decrescimento de dados.

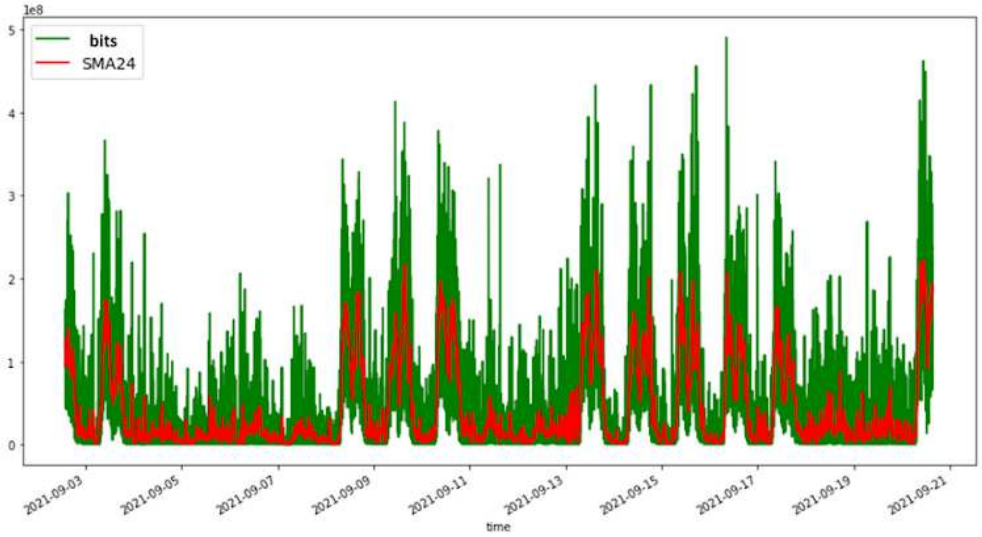
Gráfico 3: Média Móvel de Tráfego In PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

O Gráfico 3 descreve melhor o comportamento da rede nos momentos de inserção de dados, onde time é o tempo em questão com as respectivas datas de ocorrência, bits se trata da quantidade computacional dos dados e SMA24 se trata das médias móveis encontradas durante o processo de manipulação.

Gráfico 4: Média Móvel de Tráfego Out PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

O Gráfico 4 descreve melhor o comportamento da rede nos momentos de consulta de dados, onde time é o tempo em questão com as respectivas datas de ocorrência, bits se trata da quantidade computacional dos dados e SMA24 se trata das médias móveis encontradas durante o processo de manipulação.

Agora que é possível entender o comportamento, a previsão de dados também poderá ser feita.

Com o uso do Teste de Dickey-Fuller é verificado se a amostra não possui raízes unitárias, sendo assim estacionária, ou seja, que pode passar pelo processo de previsão de dados.

Tabela 1 - Resultados do Teste Dickey-Fuller em Tráfego In PACS01 to CORE02

Teste ADF	-7.468.389.229.956.250
Valor de P	0,00005131728
Quantidade da amostra	49942
Valor crítico (1%)	-3.430.480.944.618.590
Valor crítico (5%)	-28.615.978.748.306.900
Valor crítico (10%)	-2.566.800.804.858.540

A Tabela 1 evidencia que a amostra pode passar pelo processo de previsão. O "Valor de P" é menor do que 0.05, logo a série é estacionária. Isso significa que é possível manipular a amostra utilizando previsões.

Tabela 2 - Resultados do Teste Dickey-Fuller em Tráfego Out PACS01 to CORE02

Teste ADF	-7.468.389.229.956.250
Valor de P	0,00005131728
Quantidade da amostra	49942
Valor crítico (1%)	-3.430.480.944.618.590
Valor crítico (5%)	-28.615.978.748.306.900
Valor crítico (10%)	-2.566.800.804.858.540

A Tabela 2 evidencia que a amostra pode passar pelo processo de previsão. O "Valor de P" é menor do que 0.05, logo a série é estacionária. Isso significa que é possível manipular a amostra utilizando previsões.

Após isso o modelo ARIMA entra como facilitador, pois com o melhor modelo é possível encontrar a melhor previsão. O ARIMA pode falhar em diversos modelos ou se aproximar do resultado esperado.

Para este caso, os últimos 5 dados registrados no arquivo são excluídos temporariamente. Estes dados são sobre as consultas e as inserções ocorridas na rede, onde "PACS01: Tráfego de VDX PACS01 para CORE02" e "PACS01: Saída de tráfego de VDX PACS01" são os arquivos que contém tais dados.

Primeiramente o teste para a previsão dos dados deve ser feito com a remoção de uma parte da amostra. Para isso foram retirados os 5 últimos dados registrados nas amostras "PACS01: Trafego In VDX PACS01 To CORE02" e "PACS01: Trafego Out

VDX PACS01 To CORE02" afim de verificar se com o modelo ARIMA as predições estariam quase de acordo com os dados reais.

Tabela 3 – Erro encontrado para previsão em Tráfego In PACS01 to CORE02

Original	Previsão	Erro
46500040	85474691	46%
144561248	86244125	68%
80679992	97596496	17%
50845784	101102943	50%
21446240	116359881	82%

A Tabela 3 revela o erro ocorrido no cálculo com os 5 últimos dados da quantidade bits da amostra que foram removidos, onde "Erro" mostra a porcentagem, a coluna "Original" que se refere aos dados reais removidos da amostra e, a coluna "Previsão" que se trata da quantidade de bits previstas para a amostra.

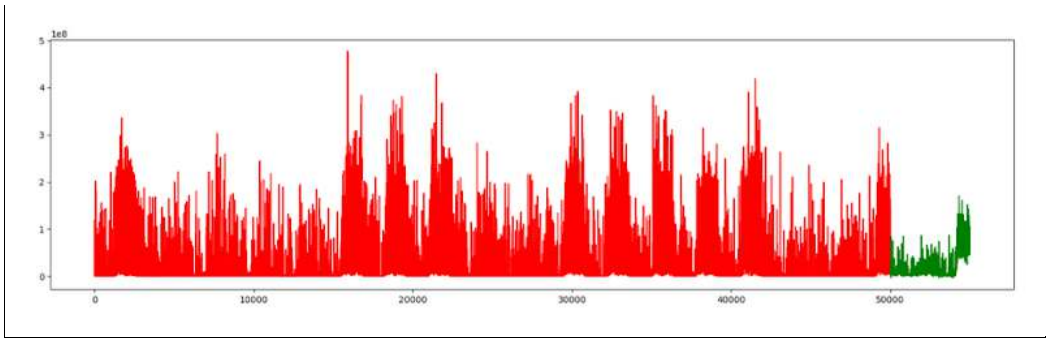
Tabela 4 - Erro encontrado para previsão em Tráfego Out PACS01 to CORE02

Original	Previsão	Erro
142768752	56481805	153%
98913568	42022125	135%
65983776	46363275	42%
79619216	36932851	116%
127371504	41771131	205%

A Tabela 4 revela o erro ocorrido no cálculo com os 5 últimos dados da quantidade bits da amostra que foram removidos, onde "Erro" mostra a porcentagem, a coluna "Original" que se refere aos dados reais removidos da amostra e, a coluna "Previsão" que se trata da quantidade de bits previstas para a amostra.

Feito isso é necessário verificar o comportamento futuro da rede. Para isso foram removidos 5mil dados dos últimos dias da amostra. Esta remoção garante um teste de previsão com menos chances de haver erros, pois quanto maior a amostra, melhor é estudado o comportamento dessa amostra.

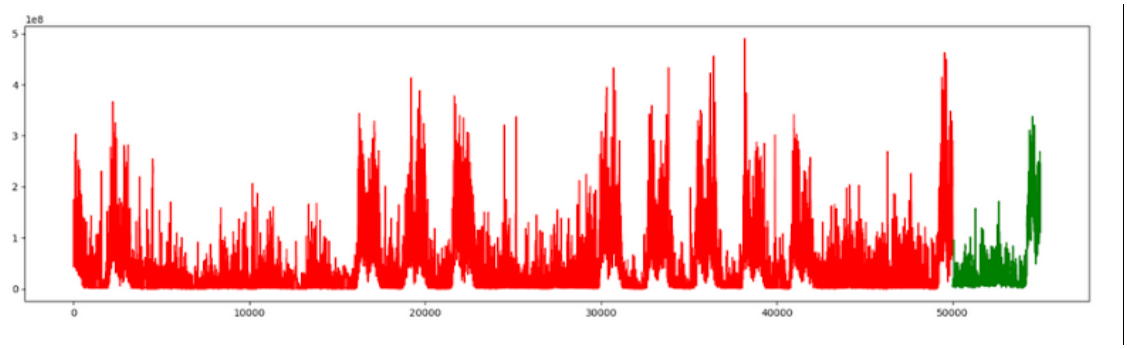
Gráfico 5: Predição com 5 mil dados em Tráfego In PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

O Gráfico 5 mostra uma predição para os próximos 5 mil dados, após o período de 2 de setembro de 2021 a 20 de setembro de 2021. O comportamento continua parecido com o comportamento dos dados reais da amostra.

Gráfico 6: Predição com 5 mil dados em Tráfego Out PACS01 to CORE02



Fonte: Elaborado pelos autores (2022)

O Gráfico 6 mostra uma predição para os próximos 5 mil dados, após o período de 2 de setembro de 2021 a 20 de setembro de 2021. O comportamento continua parecido com o comportamento dos dados reais da amostra.

CONCLUSÕES

Embora a amostra usada seja pequena, os processos feitos neste trabalho ainda podem ser usados para futuras amostras, pois todos os modelos matemáticos e computacionais serão os mesmos.

Dessa forma para grandes ou pequenos volumes de dados será possível calcular e estimar comportamentos futuros desta rede.

Por fim, este trabalho pode servir como um auxiliador em análises estatísticas de qualquer hospital, podendo se adaptar de acordo com as necessidades.

RECOMENDAÇÕES

Com o uso da linguagem de programação e as ferramentas estatísticas conhecidas, é possível extrair, visualizar e manipular qualquer tipo de dado.

No geral, com esta pesquisa torna-se possível aprender e ajudar a prever dados, assim, entendendo o comportamento futuro do fluxo de dados de uma rede.

A importância e os objetivos da pesquisa para ajudar no fluxo de dados da rede possibilita uma melhor compreensão do tráfego.

Mesmo com alguns recursos limitados, como o processamento de computadores e a pequena quantidade na base de dados, acredita-se que com novas evoluções de máquinas, será cada vez mais fácil fazer predições com análises estatísticas e computacionais.

REFERÊNCIAS

COMER, D. E.4ª (Ed.), 2007. Redes de Computadores e Internet. Bookman.

MACEDO, E. L. C. (2015). PREVISÃO DE TRÁFEGO EM ENLACES DE REDES UTILIZANDO SÉRIES TEMPORAIS, Universidade Federal do Rio de Janeiro.

PIANYKH, O. S. 2ª (Ed.), 2012. Digital Imaging and Communications in Medicine (DICOM). Springer, Berlin, Heidelberg.

STRICKLAND, N. (2000a). PACS (picture archiving and communication systems): filmless radiology, BMJ Publishing Group.

YOO, TERRY. (2004). Insight Into Images Principles and Practice for Segmentation, Registration and Image Analysis. 10.1201/b10657.