**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DE**

**SÃO PAULO**

**CAMPUS JACAREÍ**

**CURSO DE GRADUAÇÃO EM TECNOLOGIA EM ANÁLISE**

**E DESENVOLVIMENTO DE SISTEMAS**

Tamara Raquel da Silva

**RELATÓRIO DE ESTÁGIO SUPERVISIONADO I**

Jacareí-SP

Dezembro/2022

Tamara Raquel da Silva

**RELATÓRIO DE ESTÁGIO SUPERVISIONADO I**

Relatório Final apresentado como conclusão do Estágio Supervisionado I do Curso de Graduação em Tecnologia em Análise e Desenvolvimento de Sistemas, do Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Campus Jacareí, como pré-requisito para obtenção do título de Tecnólogo em Análise e Desenvolvimento de Sistemas.

Jacareí-SP

Dezembro/2022

**APRESENTAÇÃO**

Neste relatório apresento o projeto trabalhado, as atividades desenvolvidas, aprendizados e experiências adquiridas durante o período de estágio, realizado ao longo de 2021.

O estágio é continuidade de um projeto de iniciação científica, iniciado em agosto de 2020, que teve, entre outros objetivos, o de consolidar os dados utilizados como base deste projeto.

O meu trabalho consistiu em propor uma solução eficiente para resolver a questão dos dados faltantes presentes nas leituras de sensores das sondas espaciais Helios 1 e 2. Embora a solução desenvolvida tenha sido desenhada para esse conjunto de dados, as técnicas utilizadas são aplicáveis a problemas semelhantes, o que tornou esse aprendizado de grande valor para mim, como profissional.

O relatório aborda as duas fases de trabalho seguidas. A primeira fase, inicia-se com o processo que segui no estudo do problema, através de levantamento de referências. Depois, o estudo da ferramenta low-code de desenvolvimento, Orange Data Mining, sugerida pelos orientadores, em que foi desenvolvida uma solução inicial para o problema. Então, descrevo os processos de importação, visualização, manipulação, seleção e análise dos dados, utilizando linguagem de programação Python. Passo, a seguir, para as etapas de configuração dos modelos de aprendizado de máquina preditivos, treino, validação, teste e avaliação dos resultados dos modelos. Os resultados obtidos mostraram que as redes neurais são modelos adequados para a solução do problema, no entanto o software utilizado apresentou algumas limitações, de forma que os orientadores sugeriram que eu aprendesse a programar redes neurais em Python, o que me daria uma maior liberdade para configurar o modelo, com o objetivo de se chegar a um modelo que apresentasse menos erros de previsão. Na segunda fase, me dediquei a estudar de forma mais aprofundada a programação em linguagem Python e cobrindo as principais bibliotecas voltadas para aprendizado de máquina, aprendizado profundo e ciência de dados. Passei então à análise exploratória dos dados, em que utilizei as bibliotecas Python NumPy, Pandas e MatplotLib, além do Microsoft Excel para planilhas de apoio e controle. Depois disso, criei uma rotina pré-processamento dos dados. Aí iniciei uma pesquisa de trabalhos acadêmicos que abordam a temática de modelos de previsão de séries temporais, estudei, refinei a pesquisa para trabalhos que utilizavam modelos de aprendizado de máquina e, por fim, procurei por trabalhos que utilizavam dados de sensores. Após ter essas referências, iniciei a programação do modelo de rede neural do tipo recorrente Long-Short Term Memory (LSTM), considerado na literatura como adequado para a previsão de séries temporais. A programação utilizou as bibliotecas Python Keras e TensorFlow. Foram utilizados nove modelos com diferentes configurações. Outros parâmetros foram configurados também, como por exemplo, a quantidade de leituras passadas utilizadas para prever a próxima. Por fim, elaborei o treinamento, validação e teste destes modelos e avaliei os resultados obtidos.

Todos estes processos foram realizados em horário contrário ao das aulas, com reuniões semanais com os orientadores feitas de forma remota, através do software de teleconferências Microsoft Teams, devido às restrições por causa da pandemia de Covid-19.

**SUMÁRIO**

|  |  |
| --- | --- |
| **Apresentação** | **1** |
| **Identificação do campo do estágio** | **4** |
| **Introdução** | **5** |
| **Objetivos** | **6** |
| **Descrição da área do estágio** | **7** |
| **Correlação com o curso** | **8** |
| **Atividades desenvolvidas** | **10** |
| **Considerações finais** | **17** |
|  |  |

**1 IDENTIFICAÇÃO DO CAMPO DE ESTÁGIO**

1. **Identificação do local de estágio**

Nome do estabelecimento: Instituto Federal de Educação, Ciência e Tecnologia de São Paulo (IFSP) - Câmpus Jacareí

Bairro: Jardim Panorama

CEP: 12322-030

Endereço: Rua Antonio Fogaça de Almeida, 200

Cidade: Jacareí - SP

Telefone: (12) 2128 5200

Área onde foi realizado o estágio: Ensino

Data de início: 01/03/2021

Data de término: 30/11/2021

Duração em horas: 720

Supervisor do estágio: Aline de Lucas

Descrição sobre o local de estágio: Dentro da estrutura do IFSP-Jacareí existe a Coordenadoria de Pesquisa e Inovação (CPI), responsável, entre outras competências, por incentivar, apoiar, mapear e supervisionar os projetos de pesquisa e de inovação do campus. A CPI é responsável pelo processo de seleção de projetos do Programa Institucional de Bolsas de Iniciação Científica e Tecnológica do IFSP (PIBIFSP), que é realizado no segundo semestre de cada ano e regido por edital específico. O edital é sempre amplamente divulgado aos docentes e discentes. Quando o projeto é contemplado, docente e aluno têm nove meses para desenvolver as atividades previstas e entregar relatórios parcial e final. Além disso, é obrigatória a apresentação de trabalho relacionado ao projeto em evento científico/tecnológico nacional ou internacional pelo aluno. Este projeto foi contemplado no Edital nº 371/2020.

**2 INTRODUÇÃO**

Este relatório descreve as atividades referentes ao período de estágio curricular obrigatório do Curso Superior de Tecnologia em Análise e Desenvolvimento de Sistemas, oferecido pelo Instituto Federal de Educação, Ciência e Tecnologia de São Paulo, Câmpus Jacareí. O estágio teve início em 1º de março de 2021 e término em 30 de novembro de 2021. Foi realizado de forma remota, com reuniões semanais agendadas com os orientadores, Prof. Dr. Aline de Lucas, Prof. Dr. Ana Paula Shiguemori e Prof. Dr. Elcio Shiguemori.

Em 2020 tive a oportunidade de trabalhar em um projeto de Iniciação Científica relacionado ao tema do presente trabalho, em virtude da desistência do aluno que desenvolvia o projeto. Então, em 2021, a professora Aline me convidou para trabalhar neste novo projeto. Como eu tinha interesse na pesquisa e um pouco de familiaridade com os dados, fiquei contente com essa possibilidade. Além disso, o projeto exigia que eu buscasse um conhecimento mais aprofundado sobre certos temas no campo da Inteligência Artificial que não são cobertos no curso regular, o que me traria um ganho de qualificação como profissional, experiência, além do desenvolvimento de soft skills, como autonomia, responsabilidade, senso crítico, capacidade analítica e pró-atividade.

O objetivo do trabalho foi desenvolver uma solução para o problema da ausência de dados em leituras registradas por sensores a bordo das sondas Helios 1 e 2. As falhas dos sensores decorrem de vários fatores, como a interferência do clima espacial nos instrumentos, por exemplo. Existe uma quantidade muito grande de dados, por esse motivo existe a necessidade do uso de ferramentas tecnológicas para manipular esses dados e assim compreendê-los melhor.

Com a orientação dos professores, desenvolvi modelos capazes de prever os dados e, assim, reconstruir a série temporal, utilizando técnicas de aprendizado profundo (deep learning) e linguagem de programação Python. Além dos modelos, construí algoritmos de apoio, também utilizando Python.

**3 OBJETIVOS**

O principal objetivo proposto para mim era conseguir prever os períodos em que não houve leitura dos instrumentos, através da aplicação de aprendizado de máquina a séries temporais. Além disso, eu tinha os seguintes objetivos a cumprir no decorrer do período do estágio:

* Revisar a bibliografia base para entender o cenário e o problema proposto no trabalho;
* Estudar o funcionamento do software Orange, utilizado para o desenvolvimento dos modelos de aprendizado de máquina;
* Desenvolver Redes Neurais Artificiais (RNA) para previsão, utilizando o software Orange;
* Avaliar e comparar o desempenho das RNAs com outros modelos de aprendizado de máquina;
* Estudar as bibliotecas de aprendizado de máquina na linguagem de programação Python, para o desenvolvimento de modelos de previsão;
* Aprofundar meu entendimento sobre Redes Neurais Artificiais;
* Desenvolver modelos preditivos utilizando Python;
* Reconstruir os períodos sem leitura (gaps) com uso de modelo de aprendizado de máquina;
* Construir os gráficos com os dados dos eventos de choque.

**4 DESCRIÇÃO DA ÁREA DO ESTÁGIO**

O estágio abrangeu as áreas de Desenvolvimento, Inteligência Computacional e Análise de Dados, predominantemente.

Com relação ao Desenvolvimento, atuei na criação de algoritmos para leitura, consolidação, manipulação, seleção e pré-processamento dos dados, utilizando linguagem de programação Python, as IDEs Google Colaboratory e PyCharm. Além da criação de algoritmos, atuei na configuração de ambientes de desenvolvimento.

Na área de Inteligência Computacional, configurei, avaliei e testei diferentes modelos preditivos, utilizando inicialmente, técnicas de aprendizado de máquina e, posteriormente, técnicas de aprendizado profundo. Na primeira fase do estágio utilizei o software Orange Data Mining para realizar esses processo e na segunda fase utilizei a linguagem Python e a IDE Google Colaboratory.

Na área de Análise de Dados, criei rotinas para a transformação dos dados de forma a melhorar a visualização e o entendimento dos dados. Criei gráficos, tanto na fase inicial, de análise dos dados para compreender melhor a base, identificar padrões, tendências, correlações, como depois, para visualização das previsões dos modelos.

**5 CORRELAÇÃO COM O CURSO**

Durante o desenvolvimento tive a oportunidade de aplicar de forma prática conceitos aprendidos em diversas disciplinas do curso de Tecnologia em Análise e Desenvolvimento de Sistemas até o 4º período do curso. A exemplo, cito:

* **Algoritmos e Linguagem de Programação:** durante a programação de algoritmos, em diversas etapas do projeto e na compreensão de código-base de outros autores, durante o estudo de referências.
* **Estruturas de Dados:** em diversos momentos foi necessário pensar qual era a estrutura de dados mais adequada para determinada função. Além disso, nas etapas de transformação de dados, frequentemente havia a mudança da estrutura de dados, o que exigiu a compreensão das diferentes estruturas e como manipulá-las.
* **Programação Orientada a Objetos:** muitas funções existentes dentro das bibliotecas Python e que foram utilizadas no decorrer do trabalho, são orientadas a objetos.
* **Cálculo:** no entendimento do funcionamento das redes neurais artificiais, por exemplo, como se dá o cálculo dos pesos, como funciona a função de ativação, a função de custo, a função de perda, a retropropagação, entre outros.
* **Estatística:** na explicação dos resultados, cálculo e interpretação das métricas de desempenho dos modelos, entre outros.
* **Introdução à Inteligência Computacional:** no entendimento, manipulação e visualização dos dados, compreensão dos conceitos básicos relacionados à Inteligência Artificial, no desenvolvimento e avaliação dos modelos.
* **Metodologia de Pesquisa Científica e Tecnológica:** na estruturação de todo o trabalho, definição da metodologia seguida e no levantamento de referências.
* **Inglês Técnico:** na compreensão de trabalhos de referência redigidos em língua inglesa, na leitura dos manuais das bibliotecas Python, em vídeos tutoriais de programação, em sites de consulta e na redação do relatório final e resumo expandido.
* **Comunicação e Expressão:** na escrita dos relatórios, resumo expandido e elaboração de apresentações.

**6 ATIVIDADES DESENVOLVIDAS**

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Revisão bibliográfica.

**Por que foi feito:** Para que eu pudesse entender melhor o problema existente, o que me ajudaria a compreender melhor a base de dados e estruturar a solução.

**Como foi feito:** A professora Aline disponibilizou uma coletânea de trabalhos relacionados ao problema, eu tive um prazo para a leitura e depois foram realizadas reuniões em que ela explicou melhor o problema, discutiu o assunto comigo e respondeu as dúvidas que eu tinha.

**Qual a aprendizagem com a atividade:** Além de melhorar a minha compreensão do problema, que era essencial para o trabalho, foi bom para aprender melhor como estruturar e apresentar um trabalho científico. Também foi importante aprender esse processo de busca do conhecimento, pois na vida profissional nem sempre eu vou ter domínio da área do conhecimento para a qual vou desenvolver as soluções tecnológicas e será necessário seguir esse processo. Além disso, essa atividade me ajudou a aprender coisas como fazer as perguntas certas, a importância do escopo do projeto, como apresentar os resultados, entre outras.

1. Locais do estágio: IFSP-Câmpus Jacareí, PGCAP/INPE, Internet

**O que foi feito:** Preparação dos dados para entrada nos modelos preditivos.

**Por que foi feito:** Para se chegar ao resultado esperado, era necessário preparar os dados antes. Os modelos exigem dados pré-processados, não dados brutos.

**Como foi feito:** Eu tinha disponível uma base de dados que consolidava as leituras das duas sondas (Helios 1 e 2), no período de 1974 a 1985 (correspondente a um ciclo solar de 11 anos). Essa base foi construída no projeto que participei em 2020. Os dados foram analisados juntamente com os orientadores e foi escolhido um parâmetro, a velocidade dos prótons (Vp), para ter seus dados processados e servir de entrada para os modelos preditivos. Construí um algoritmo para fazer a seleção desses dados e identificar os gaps na série temporal. Então fui orientada a identificar e selecionar períodos da série temporal em que haviam doze leituras completas consecutivas, ou seja, sem gaps. Programei uma rotina para fazer esse procedimento de forma automática. Em seguida, foi desenvolvido um algoritmo que reorganizava os dados em uma matriz, selecionando da primeira leitura à décima segunda, na próxima linha, da segunda à décima terceira e assim por diante. Dessa forma, o modelo utilizaria doze leituras passadas para prever uma (a próxima).

**Qual a aprendizagem com a atividade:** Aprendi que a preparação dos dados é uma etapa importante para a boa performance do modelo.

1. Locais do estágio: IFSP-Câmpus Jacareí, PGCAP/INPE, Internet

**O que foi feito:** Estudo da utilização do software Orange Data Mining, aplicado ao aprendizado de máquina.

**Por que foi feito:** O software Orange Data Mining trata-se de uma ferramenta low-code, ou seja, não requer o conhecimento de uma linguagem de programação específica para ser utilizado, mas baseia-se na lógica de programação em seus componentes - os componentes são blocos que executam funções pré-programadas. Esse software permite a modelagem, treinamento, validação, teste a avaliação dos modelos de aprendizado de máquina, foi sugerido o seu uso pelos orientadores, em um primeiro momento, a título de avaliação do desempenho de modelos de redes neurais artificiais preditivas para o conjunto de dados escolhido. Como eu não tinha familiaridade com essa ferramenta, foi necessário que eu a estudasse para poder desenvolver o trabalho.

**Como foi feito:** O estudo da utilização do Orange se deu, inicialmente, pela pesquisa de vídeos tutoriais na internet. Depois, li a documentação do software que explica como utilizar alguns recursos. Dentro do próprio software são disponibilizados alguns exemplos, que reproduzi. A professora Ana Paula disponibilizou também um exemplo, para que eu executasse e depois tirou algumas dúvidas sobre a utilização do software. Além disso, participei de algumas aulas da pós-graduação em Computação Aplicada do INPE, como ouvinte, a convite do professor Elcio, em que ele apresentou possibilidades de uso do Orange. Já no segundo semestre do ano, passei a ter aulas da disciplina Introdução à Inteligência Computacional no curso regular, em que aprendi também sobre a utilização deste software.

**Qual a aprendizagem com a atividade:** Aprendi sobre a importância de se ler a documentação de um sistema e como utilizá-lo. Mais do que isso, aprendi que é fundamental se fazer um sistema bem documentado, de forma a não dificultar o seu uso pelos mais diversos públicos. Também compreendi que o fato de um software ser low-code não significa que seja de fácil utilização ou que não exija a compreensão de conhecimentos fundamentais de lógica de programação e dos algoritmos por trás dos modelos utilizados.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Desenvolvimento das Redes Neurais Artificiais (RNAs), utilizando o software Orange.

**Por que foi feito:** Essa atividade foi desenvolvida com o objetivo de entender se as redes neurais seriam bons modelos para previsão do conjunto de dados escolhido.

**Como foi feito:** Primeiramente fiz a importação dos dados pré-processados no Orange e separei a amostra em conjunto de treinamento e conjunto de teste. O conjunto de treinamento consiste em dados rotulados que serão utilizados pelo modelo para aprender padrões da série temporal, enquanto o conjunto de teste é formado por dados que o modelo não viu - a amostra foi dividida em 70% para treino e 30% para teste. Depois, fiz a configuração dos hiperparâmetros das redes neurais. Então realizei o treinamento e em seguida o teste dos modelos.

**Qual a aprendizagem com a atividade:** Com esta atividade aprendi de forma prática como funcionam os modelos de redes neurais, além de como configurar treinos e testes dos modelos.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Desenvolvimento de outros modelos preditivos utilizando o software Orange.

**Por que foi feito:** Essa atividade foi desenvolvida com o objetivo de comparar a performance das redes neurais com outros modelos.

**Como foi feito:** A parte de importação dos dados e separação da amostra em treino e teste seguiu da mesma forma que foi feita para as RNAs. Então configurei os modelos Adaboost e Random Forest. Os modelos foram treinados e testados e seus resultados obtidos.

**Qual a aprendizagem com a atividade:** Aprendi que é importante comparar os resultados de uma técnica com outras.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Análise dos resultados obtidos com a utilização do Orange.

**Por que foi feito:** A análise dos resultados obtidos, além de permitir que eu entendesse melhor os resultados a partir das métricas, foi fundamental para determinar o seguimento do projeto - qual modelo seria utilizado.

**Como foi feito:** O Orange calcula as seguintes métricas de avaliação para os modelos preditivos: Erro Quadrático Médio (MSE), Raiz do Erro Quadrático Médio (RMSE), Erro Absoluto Médio (MAE) e Coeficiente de Determinação (R2). Com base nas métricas e nas previsões, elaborei um relatório comparando os resultados dos modelos. As RNAs apresentaram o menor erro de previsão, quando comparadas com os demais modelos. Além do relatório, foi feita uma apresentação dos resultados preliminares para os orientadores e no congresso do IFSP, Conict.

**Qual a aprendizagem com a atividade:** Para a realização dessa atividade foi necessário revisar conceitos estatísticos para entender e explicar as métricas obtidas. Além disso, aprendi a estruturar um relatório de avaliação dos modelos e a apresentação dos resultados.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Análise exploratória dos dados com Python.

**Por que foi feito:** Essa atividade foi feita para analisar o comportamento dos dados, das variáveis, identificar padrões, verificar a quantidade de valores faltantes, comparar os dados da série ao longo dos anos, entre outros motivos.

**Como foi feito:** A análise do conjunto de dados foi feita com o auxílio de um notebook, usando o ambiente de programação Google Colaboratory e linguagem de programação Python. Foram utilizadas as bibliotecas Numpy, Pandas e MatplotLib. Também fiz alguns gráficos, por exemplo, mostrando os dados de um mesmo período ao longo dos anos, para entender melhor a amostra.

**Qual a aprendizagem com a atividade:** Aprendi que as bibliotecas Python podem oferecer diversos tipos de solução para visualização e análise de dados em um único ambiente, o que facilitou bastante a minha análise. Sem o uso dessa ferramenta tecnológica teria que utilizar diferentes ferramentas desintegradas, o que poderia gerar perdas de eficiência. Também aprendi alguns padrões acerca do conjunto de dados utilizado, o que me ajudou a compreender melhor os resultados finais.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Pré-processamento dos dados em Python.

**Por que foi feito:** Essa atividade foi necessária para a adequação dos dados para servirem de entrada para o modelo de rede neural recorrente.

**Como foi feito:** Foi desenvolvido um código para pré-processar os dados, de forma que estivessem adequados para seu uso nas RNAs. Foi feito o filtro dos valores NaN (not a number), que representam as falhas de medição. Além disso, o código também fazia a seleção das colunas de interesse, a normalização dos dados, entre outras funções de pré-processamento.

**Qual a aprendizagem com a atividade:** Aprendi novas formas, funções e algoritmos que podem ser utilizados para pré-processar dados.

1. Local do estágio: IFSP-Câmpus Jacareí

**O que foi feito:** Escolha e construção do modelo em Python.

**Por que foi feito:** Em Python eu teria uma maior liberdade para alterar configurações das redes neurais e utilizar modelos mais adequados para a previsão desse conjunto de dados do que o modelo de RNA multilayer perceptron, utilizado no Orange.

**Como foi feito:** O modelo escolhido foi de uma rede neural recorrente e profunda LongShort Term Memory (LSTM). A escolha se deu por ser um dos modelos considerados na literatura como adequados na aplicação de previsão de séries temporais. A diferença de uma RNA unidirecional para uma RNA LSTM, de forma resumida, se dá pelo fato de que os neurônios na LSTM são capazes de fazer mais conexões, além disso, as LSTM possuem uma célula de memória, o que contribui para a resolução de problemas de maior complexidade. Foi configurado um modelo inicial e seguidas as etapas de compilação, treinamento, validação, teste e avaliação das métricas.

**Qual a aprendizagem com a atividade:** Aprendi como funciona um modelo de RNA recorrente. Também aprendi a programar e configurar o modelo e as etapas de compilação, treinamento, validação e teste.

1. Local do estágio: PGCAP/INPE

**O que foi feito:** Participação como ouvinte nas aulas da disciplina Inteligência Artificial.

**Por que foi feito:** Essa atividade foi desenvolvida para que eu pudesse compreender melhor as bases de Inteligência Artificial e os conceitos existentes por trás dos modelos utilizados neste projeto.

**Como foi feito:** As aulas fazem parte do programa de pós-graduação em Computação Aplicada, oferecido pelo Instituto Nacional de Pesquisas Espaciais (INPE). O convite foi feito pelo professor colaborador neste projeto, Dr. Elcio Hideiti Shiguemori. Foi feita uma revisão do processo a ser seguido na elaboração do projeto, em que foram explicadas cada uma das etapas: Escolha do problema, escolha da técnica, obtenção dos dados, pré-processamento dos dados, aplicação das técnicas e análise dos resultados. Depois, foram estudados os fundamentos de aprendizagem de máquina, incluindo a abordagem de redes recorrentes e a questão da importância do treinamento bem feito. Então, foi apresentado o desenvolvimento de modelos utilizando-se o software Orange, levantando-se pontos bastante pertinentes, por exemplo: como lidar com dados faltantes no conjunto de dados; como verificar se o modelo está especializado; como testar o modelo com dados novos, entre outros. Foram discutidas também as limitações do software Orange, como: só aceita uma saída; não permite salvar épocas de treinamento; requer a interação; e, não apresenta todos os tipos de arquitetura de redes neurais.

**Qual a aprendizagem com a atividade:** Aprendi diversos conceitos necessários ao desenvolvimento das atividades do estágio. Também aprendi melhor sobre a utilização e as limitações do software Orange.

1. Local do estágio: Ambiente de aprendizagem da Alura.

**O que foi feito:** Realizei o curso Deep Learning: Previsão com Keras.

**Por que foi feito:** Essa atividade me ajudou na parte prática da programação do modelo em Python, utilizando a biblioteca Keras, além de revisar importantes conceitos na área de aprendizado profundo.

**Como foi feito:** O curso foi desenvolvido de forma online, com aulas teóricas e práticas e questões sobre o conteúdo apresentado a serem respondidas. Também foi desenvolvido um projeto de aplicação.

**Qual a aprendizagem com a atividade:** Aprendi de forma prática a programar o modelo de rede neural utilizando Python, além de questões relacionadas à configuração de ambiente.

1. Local do estágio: Diversos

**O que foi feito:** Assisti a palestras disponibilizadas de forma online durante diversos congressos.

**Por que foi feito:** O objetivo foi melhorar o meu desenvolvimento e a minha capacidade de resolução de problemas relacionados ao projeto.

**Como foi feito:** Durante o ano de 2021, em virtude da pandemia de Covid-19, várias universidades divulgaram seus congressos na modalidade remota e aberta à comunidade. Dessa forma, foi uma grande oportunidade para aprender mais sobre os temas relacionados ao projeto. A participação foi feita via inscrição nos sites dos eventos, o link para assistir aos eventos foi enviado para mim e depois eu recebi os certificados de participação.

**Qual a aprendizagem com a atividade:** A participação em palestras representa uma oportunidade para que eu me desenvolvesse mais, aprendesse sobre assuntos relacionados ao estágio e pudesse ter um maior contato com o ambiente de pesquisa científica, podendo observar e aprender mais sobre os diferentes métodos de pesquisa utilizados. Cabe acrescentar, que além de me ajudar nas atividades do estágio, as participações contribuíram também para minha formação profissional de forma complementar.

**10 CONSIDERAÇÕES FINAIS**

A oportunidade de poder participar desse estágio através do Programa de Iniciação Científica do IFSP foi um grande presente. Neste estágio pude ter contato com profissionais de diversas áreas, experientes, capazes, pacientes e que me ajudaram sempre que eu precisei de orientação. Também foi uma porta que se abriu para mim na área da pesquisa e pude aprender muito, pude ter contato com trabalhos no estado da arte da tecnologia e ver o desenvolvimento de novas técnicas. Vale ressaltar que grande parte do que aprendi nesse estágio vai além do currículo do curso de Análise e Desenvolvimento de Sistemas, então valorizo bastante essa experiência.

Além disso, o estágio proporcionou um tempo para que eu pudesse aprender mais sobre as disciplinas que foram cursadas e deu oportunidades para que eu pudesse aplicar os conhecimentos adquiridos de forma prática, voltada à resolução de um problema complexo da vida real.

Por fim, a experiência do estágio exigiu que eu me desenvolvesse como pessoa e como profissional, aprendendo a: lidar com demandas complexas; ouvir os conselhos, instruções, sugestões, críticas e correções dos orientadores; buscar referências diversas; ser responsável com o conteúdo desenvolvido, com a técnica, o método, os prazos; fazer as perguntas corretas; ter autonomia para trabalhar e, quando necessário, tomar decisões; desenvolver meu senso crítico e capacidade de análise; entender o outro e me fazer entender; traduzir conceitos complexos; entender e explicar os limites da tecnologia, entre outras coisas.

Sou muito agradecida por este período de estágio e acredito que este tipo de experiência agregou muito à minha vida como pessoa, como estudante e como futura profissional na área de tecnologia.

**REFERÊNCIAS**

Relação dos autores e obras consultadas no decorrer das atividades desenvolvidas.

ALBON, C. Machine Learning with Python Cookbook: Practical Solutions from Preprocessing to Deep Learning. O’Reilly Media, 2018.

BITARÃES, S. M. Previsão de séries temporais em processos de mineração utilizando Redes Neurais Recorrentes e séries temporais nebulosas. Ouro Preto: UFOP, 2019.

CANE, H. V.; KAHLER, S. W.; SHEELEY JR., N. R. Interplanetary shocks preceded by solar filament eruptions. J. Geophys. Res., v. 91, p. 13321–13329, Dec. 1986.

CHARNIAK, E. Introduction to Deep Learning. The MIT Press, 2019.

DATA SCIENCE ACADEMY. Deep Learning Book. Disponível em: < [https://www.deeplearningbook.com.br/](about:blank)>.

GÉRON, AURÉLIEN. Mãos à obra: Aprendizado de Máquina com ScikitLearn & TensorFlow. Rio de Janeiro: Alta Books, 2019.

GONZALEZ, W. et al. Interplanetary origin of geomagnetic storms. Space Sci. Rev., v. 88, p.529–562, 1999.

HAYKIN, S. Redes Neurais: Princípios e práticas. 2ª ed. Bookman, 2005.

HUTTUNEN, K. E. J.; KOSKINEN, H. E. J. Importance of post-shock streams and sheath region as drivers of intense magnetospheric storms and high-latitude activity. Ann. Geophys., v. 22, p.1729–1738, 2004.

KERAS v. 2.4.0. Disponível em: < https://keras.io/ >.

MARSCH, E.; SCHWENN, R. Introduction. In: SCHWENN, R.; MARSCH, E. (Ed.). Physics of the inner heliosphere. Berlin, Germany; New York, U.S.A.: Springer, 1990. p.1–12.

MULLER, A.; GUIDO, S. Introduction to Machine Learning with Python: A guide for data scientists. O’Reilly, 2016.

NUMPY v.1.19.0. Disponível em: < https://numpy.org/ >.

ORANGE. Disponível em: <https://orange.biolab.si>.

PANDAS v. 1.1.5. Disponível em: <https://pandas.pydata.org>.

PORSCHE, H. General aspects of the mission Helios 1 and 2. Introduction to a special issue on initial scientific results of the Helios Mission. Journal of Geophysics Zeitschrift Geophysik, v. 42, p.551–559, 1977. 16

PORSCHE, H. Ten years of HELIOS. Oberpfaffenhoven: NASA STI/Recon Technical Report N, 1984. 27794-+ p.

PYTHON. Disponível em: <https://python.org.br>.

RASCHKA, S.; MIRJALILI, V. Python Machine Learning: Machine Learning and Deep Learning with Python, scikit-learn, and TensorFlow. Packt, 2017.

ROSENBAUER, H. et al. A survey on initial results of the HELIOS plasma experiment. Journal of Geophysics, v. 42, p.561-580, 1977.

RUSSELL, C. T. et al. The interplanetary shock of September 24, 1998: Arrival at Earth. J. Geophys. Res., v. 105, p.25143–25154, nov 2000.

SCHWENN, R. An essay on terminology, myths and known facts: solar transient - flare - cme - driver gas - piston - bde - magnetic cloud - shock wave - geomagnetic storm. Astrophys. Space Sci., v. 243, p. 187, 1996.

SCHWENN, R. Space Weather: The Solar Perspective, Living Rev. Sol. Phys., 2006. Acesso em: 27 de setembro de 2020. Disponível em: .

SCIKIT-LEARN v.0.23.2. Disponível em: <https://scikit-learn.org>.

SHEELEY JR., N.; HOWARD, R. A.; MICHELS, D. J.; KOOMEN, M. J.; SCHWENN, R.; MUHLHAUSER, K. H.; ROSENBAUER, H. Coronal mass ejections and interplanetary shocks. J. Geophys. Res., v. 90, p.163–175, 1985.

SHIGUEMORI, E. H. Aprendizagem de máquina, Redes Neurais Artificiais, aprendizado supervisionado, não supervisionado, por reforço. CAP/INPE: São Paulo, 2021a.

SHIGUEMORI, E. H. Preparação dos dados, ferramentas e bibliotecas. CAP/INPE: São Paulo, 2021b.

SHIGUEMORI, E. H. Bibliotecas, neurônio artificial, tipologia, funções de ativação, tipos de aprendizagem. CAP/INPE: São Paulo, 2021c.

SHIGUEMORI, E. H. Aprendizado Profundo, Redes Profundas, Redes Convolucionais, Redes Recorrentes, Redes Pré-treinadas, Metaheurísticas, Algoritmos Genéticos. CAP/INPE: São Paulo, 2021d.

TENSORFLOW 2. Disponível em: < https://www.tensorflow.org/>.

TSURUTANI, B.T. et al. Origin of interplanetary southward magnetic fields responsible for major magnetic storms near solar maximum (1978–1979), J. Geophys. Res., 93, p.8519–8531, 1988.

TSURUTANI, B. et al. Great magnetics storms. Geophysical Research Letters, v. 19, p.73–76, 1992.

W3SCHOOLS. Python Tutorial. Disponível em: <https://w3schools.com.



**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Aline de Lucas

Orientadora do estágio

**\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_**

Tamara Raquel da Silva

Estagiária