

# Willian Werner

ESTAGIÁRIO · DATA SCIENCE

☎ (11) 96461-2661 | 🏠 willianw.github.io | 💻 willianw | 🔑 kwillianw | 📄 willian-werner-9759599a |

✉ willianwernerangelodacosta@gmail.com

## Formação

### E.M.E.F. Coronel Romão Gomes

São Paulo

ENSINO FUNDAMENTAL I E II

2003 - 2008

- Finalista na 2ª OBMEP
- 8º lugar no Campeonato Regional de Xadrez Individual (2007)
- Campeão do Torneio interno de Xadrez (2008)

### Colégio Integrado Objetivo

São Paulo

ENSINO FUNDAMENTAL II E MÉDIO

2009 - 2013

- Bolsista pela fundação ISMART
- Participação no Programa Objetivo de Incentivo ao Talento - Programação de Jogos
- Participação em diversas olimpíadas científicas, ganhando muitas medalhas, dentre elas ouro e participação nas seletivas internacionais da OBF, e OBA, ouro na IYPT, prata na IJSO, OBQJr, OPF, dentre outras.

### Escola Politécnica da Universidade de São Paulo

São Paulo

GRADUAÇÃO EM ENGENHARIA DE COMPUTAÇÃO

Jan. 2014 - PRESENT

- Participação pelo time de rugby da Poli na competição de calouros (2014)
- Participação na equipe de karatê da Poli (2015)
- Grupo de Teatro da Poli (2014-2017, 2019)
- Maratona de Programação - IME (2014)
- Curso de Estágio em Programação (2016)
- Participação na Hackaton USP (2017)

## Experiência

### Media Portal Soluções

São Paulo

ESTÁGIO EM FULL STACK DEVELOPMENT

mai 2016 - ago 2016

- Refatoração completa do Back-end: agregação e divisão de servlets, otimização de trajetórias, criação de um único servlet de controle
- Refatoração do banco de dados: remoção de colunas inutilizadas e agregação de colunas e tabelas, afim de eliminar redundâncias
- Remodelação geral do Front-end
- Customização da página (Front-end) para diversos clientes, remoção de bugs em clientes ativos
- Criação de web services, gerando as WDSL's correspondentes
- Tecnologias e linguagens usadas: Java, SQL, AJAX, Javascript, JSTL, JQuery, HTML, XML, CSS, Bootstrap, Ambiente Linux

## BTG Pactual

ESTÁGIO EM DEVOPS

São Paulo

jan 2017 - abr 2017

- Aprimoramento do sistema de Continuous Delivery em ambientes paralelizados, usando Blue-Green Deployment
- Aprimoramento do sistema de integração com os serviços Cloud da Amazon Web Services
- Sistema automático de controle de release para as aplicações, identificando as dependências e atualizando a versão das aplicações agrupadas
- Criação de scripts de automação de compilação, teste e distribuição das aplicações nos servidores Cloud
- Tratamento do banco de dados do Digital (Internet Banking) para o sistema de recomendação de carteira, utilizando Machine Learning
- Tecnologias e linguagens usadas: Java EE, Python, Jenkins, SQL, Git, SSH, Groovy, Ambiente Linux, Ambiente AWS

## Big Data Brasil

ESTÁGIO EM DATA SCIENCE

São Paulo

set 2017 - dez 2017

- Participação de aulas internas sobre Data Science e Machine Learning
- Desenvolvimento e aprimoramento de crawlers para captura de dados
- Criação de dashboards para visualização dos resultados por parte dos clientes
- Tecnologias e linguagens usadas: Python, R, Scrapy, Pandas, Jupyter, SQL, Power BI, Ambiente Linux, Ambiente AWS

## Mandala Clothing

ESTÁGIO EM DATA SCIENCE

São Paulo

mar 2018 - dez 2018

- Análise de perfis de clientes (clusterizing de personas)
- Análise das principais características das clientes com alto ticket médio
- Sistema de recomendação de peças
- Sistema de recomendação de tamanho de peças
- Tecnologias e linguagens usadas: Python, Pandas, Jupyter, SQL, Power BI, Ambiente Linux, Ambiente AWS, Google Sheets/Google Scripts API

## InfoSimples

ESTÁGIO EM MACHINE LEARNING

São Paulo

mai 2019 - ago 2019

- Estudo comparativo entre indexadores de texto: Elasticsearch, Azure e Sphinx, em termos de performance, custo e flexibilidade
- Protótipo para Face Verification: processamento das imagens, uso de transfer learning com o FaceNet
- Construção de uma plataforma de padronização de datasets de treino para o Recaptcha
- Tecnologias e linguagens usadas: Python, Pandas, Jupyter, SQL, Ambiente Linux, GCP

# Iniciação Científica

## Mecanismos de controle para fins didáticos

Depto. de Telecomunicações e  
Controle - Poli USP

PROF. BRUNO AUGUSTO ANGÉLICO

2015

- Três dispositivos de controle
- Comunicação do microcontrolador com o sistema externo de controle, em Simulink (Matlab)
- Ajustes dos parâmetros de comunicação serial: verificação de paridade e baud rate

## **Análise da contribuição do fator humano na confiabilidade de sistemas complexos**

*Departamento de Engenharia*

*Naval - Poli USP*

PROF. MARCELO RAMOS MARTINS

2016

- Classificação de qualidade de operadores em sistemas complexos (ex.: usinas nucleares)
- Utilização de redes neurais bayesianas para identificação das qualidades dos operadores
- Aquisição e tratamento dos dados em Matlab

## **Cursos**

---

### **Inferential Statistics**

*Duke University*

COURSERA

- Uso de métodos de inferência estatística para dados numéricos e categóricos
- Teste de hipóteses e p-values
- Interpretabilidade para publicação de resultados
- Utilização da linguagem R

### **Machine Learning**

*Stanford University*

COURSERA

- Aprendizagem supervisionada (algoritmos paramétricos e não paramétricos, máquinas de vetores suporte, núcleos e redes neurais)
- Aprendizagem não supervisionada (Algoritmos de agrupamento - clustering, redução dimensional, sistemas de recomendação, a algoritmos hierárquicos)
- As melhores técnicas de aprendizagem automática (viés indutivo / variância)
- Processo de inovação em aprendizagem automática e inteligência artificial

### **How to win a Data Science Competition**

*National Research University Higher*

*School of Economics*

COURSERA

- Como resolver problemas de competições de modelagem preditiva de forma eficiente
- Entender como essas habilidades podem ser aplicáveis em tarefas do mundo real
- Como preprocessar os dados e gerar novas features de diversas fontes como textos e imagens
- Técnicas avançadas de Feature Engineering, e.g., mean-encodings, usando medidas estatísticas agregadas ou encontrando os 'Vizinhos mais próximos' para melhorar as predições
- Metodologias confiáveis de validação cruzada, afim de metrificar modelos e evitar sobreajuste/subajuste
- Exemplos práticos de análise e interpretação dos dados. Verificação de inconsistências, ruídos, erros e outros problemas relacionados aos dados, como vazamentos; e como resolvê-los
- Exposição a diferentes algoritmos e como ajustar os hiperparâmetros de forma eficiente
- Como combinar diferentes modelos de aprendizado de máquina
- Análise completa de soluções vencedoras de competições

CS224N

- Representações de palavras em espaços vetoriais: One-Hot, Word2Vec, métricas de eficiência de representação
- Estrutura linguística: estrutura sintática, consistência, árvores de dependência, geração de árvores sintáticas com DL
- Modelos de linguagem e RNN's: evolução de modelos n-grams, redes neurais de janela fixa, até RNN's
- Variações das RNN's: problema do 'vanishing gradient', LSTM, GRU, RNN's bidirecionais e de múltiplas camadas
- Tradução automática: métodos estatísticos, Seq2Seq, busca gulosa e em feixes de traduções alternativas, métrica BLEU de performance
- Mecanismo de atenção: problema de gargalo no Seq2Seq e variantes de atenção
- Uso de CNN's: subsequências de frases, Q-RNN's e casos de uso
- Modelos de subpalavras: modelos de caracteres, modelos híbridos, FastText
- Representações de palavras em contexto e pré-treino: transfer-learning, ELMo, ULM-fit, transformadores e BERT
- Tree-RNN's e análise de sentimento: recursividade da linguagem, redes recursivas e casos de uso

## Outros

---

### Idiomas

- Inglês: Avançado
- Alemão: Básico
- Francês: Básico

### Tecnologias e Linguagens (resumo)

- Linguagens: C, C#, C++, Clojure, Elixir, Groovy, Java, Julia, Matlab, Octave, Python, Ruby, R
- Frameworks: AWS-CLI, Keras, Pandas, Scikit-learn, Scrapy, XGBoost
- Banco de Dados: MongoDB, SQL
- Front-end: CSS, HTML, Javascript

### Projetos

Os principais registros de meus projetos estão na minha homepage.