

## EDUCATION

---

- **Rice University** Aug 2022 – Dec 2027 (expected)  
*PhD in Computer Science focused in Machine Learning and Data Mining*
- **University of São Paulo (USP), ICMC** Feb 2017 – Dec 2021  
*Bachelor's degree in Computer Science; GPA: 3.36 (8.4/10 - Top 16%)*

## WORK AND RESEARCH EXPERIENCE

---

- **Experian** Oct 2020 - June 2022  
*Data Analyst I*
  - **Machine Learning:** Development and deployment of more than 100 ML models (Logistic Regression, LGBM, XGBoost) applied to consumer profile segmentation data from 210 million Brazilians.
  - **ETL:** Creation of an internal library of pipelines to automate data preprocessing and speed-up experiments using Apache Spark and Hadoop.
- **Imaflora** Nov 2020 - Apr 2022  
*Research Internship*
  - **Graph embedding and Markov Chain:** Modeling of the Amazon Rainforest timber flux government data from 2014 to 2020 using Graph Embedding techniques and Markov Chain models to rank the reliability of wood transportation chains.
  - **Data Visualization and ETL:** Built a data visualization tool to explore the chains classified by our model and the aggregated data from each wood merchant. Created ETL pipelines with Spark to process the large amount of transactions into many smaller and more informative networks, and built Julia scripts to use parallel computing to pre-process our data in real-time in a web server.
- **GNNs and Event Embedding - Universidade de São Paulo (USP)** Oct 2020 - Oct 2021  
*Research Assistant*
  - **Deep Learning and Graph Attention Networks:** Developed GNEE, a new Event Graph Embedding technique based on Graph Attention Networks (GAT) that supports heterogeneous networks in a semi-supervised training regime. Work conducted under the supervision of Professor Ricardo Marcacini
  - **Paper and Undergraduate Thesis extension:** GNEE was accepted in ICDM 2021 (International Conference on Data Mining) as a short paper. These results were extended in my Undergraduate Thesis, which compared the performance of our proposed method in this task with other techniques such as GAT and GCN in streams of data, finding that GNEE is, in many cases, superior in Accuracy and F1 while maintaining a linear computational cost.
- **ModCovid - 19 (CeMEAI)** Apr 2020 - Jul 2020  
*Research Internship - supported by the Serrapilheira Institute*
  - **Machine Learning with Time Series:** Contributed to the development of an automatic control system to assist in designing efficient mitigation measures for the Covid-19 epidemic in cities using time series models such as ARIMA and Prophet. Acknowledged in the paper "Robot Dance".
  - **Data Visualization and ETL:** Using tools such as Selenium, Pandas and D3.js, developed an automated pipeline to scrape, preprocess and visualize the extracted data from medical forms and fed Machine Learning models and Dashboards to support different experiments.
- **Privacy Policies and Joker - Universidade de São Paulo (USP)** Jul 2018 - Jul 2020  
*Research Assistant - supported by Fapesp (Sao Paulo State Research Support Foundation)*
  - **Human Computer Interaction (HCI) and Data Mining:** Developed Joker, an Open Source Card Sorting tool to support collection the clusterization of usability datasets. The platform enables researchers to run remote experiments with a large number of participants through crowdsourcing.
  - **Results and collaborations:** Our paper in collaboration with researchers of Ontario Tech University was published in SIGDOC 2019 as a result of this research project.

## MAJOR PROJECTS

---

- **Coordinator of Data ICMC- Extracurricular Data Science Group** Jan 2019 - Dec 2021
  - **Teaching and study groups:** Taught our nationally acknowledged Introduction to Machine Learning course and supervised our Deep Learning study group, focused on state-of-the-art papers from venues such as ICML, NeurIPS, and ICLR.
  - **Project supervision:** Conducted the development of more than 10 Machine Learning and Deep Learning projects about Computer Vision, Audio, and Natural Language Processing.
  - **Events:** Organized lecture series and panels covering areas in the frontiers of Deep Learning, such as "Understanding Deep Learning" and "Open-Endedness Panel". Both had lecturers from DeepMind, Google Brain, UCL, Oxford and many other affiliations.
- **Magic: The Gathering Adversarial Network (MTGAN)** Ago 2020 - Nov 2020
  - **Dataset Collection and preprocessing:** Scrapped more than 20GB of high-resolution images from different platforms. Filtered the data by removing duplicates using various computational techniques and by cleaning corrupted images. Preprocessed the entire dataset to avoid overhead while loading data to our StyleGANv2 pipeline.
  - **Main Objective:** Understand the discrepancies between the capacities of Transfer Learning of different sized GANs (Generative Adversarial Networks), ranging from DCGAN to StyleGANv2.
- **Finalist of Interhack - Brazil's biggest hackathon for students** Jul 2019 - Dec 2019
  - **Final position:** Awarded 3rd out of 180 participants
  - **Deep Learning Speech Training Application:** Used Tensorflow.js to embed a speech recognition Deep Learning model in an educational game to assist children with cognitive disabilities to improve speech skills.

## PAPERS

---

- *Semi-Supervised Graph Attention Networks for Event Representation Learning* - **João Pedro R. Mattos** and Ricardo Marcacini, to appear in ICDM 2021
- *Smart toys and children's privacy: usable privacy policy insights from a card sorting experiment* - André de Lima Salgado, Felipe Silva Dias, **João Pedro Rodrigues Mattos**, Renata Pontin de Mattos Fortes, Patrick CK Hung - SIGDOC 2019

## TECHNICAL SKILLS

---

**Languages:** C/C++, Python, JavaScript, SQL

**DS/ML/Viz:** SciPy, Scikit-Learn, Pandas, PySpark, PyTorch, Seaborn, D3, Plotly Dash

**Others:** BeautifulSoup, Scrapy, Django, Git, Docker, NodeJS, Linux, Hadoop

## AWARDS

---

- **Experian Recognition Award:** The award is given to the most efficient Experian team in Brazil by the end of the fiscal year.
- **Finalist of Interhack:** 3rd out of 180 participants in Brazil's biggest hackathon for students.

## TEACHING

---

- **Introduction to Machine Learning - Data** Jan 2020 - Jul 2020
  - **Teaching:** Taught the Introduction to Machine Learning course to more than 120 students at USP through Data, our extracurricular Data Science and Machine Learning group. The online lectures available on YouTube have over 20.000 views, being one of the biggest Machine Learning courses in Portuguese. The course bibliography was based on the highly regarded Machine Learning and Deep Learning classes from Stanford, MIT, and Cornell University.
- **Teaching Assistant Scholarship** Jan 2018 - Jul 2018
  - **Teaching:** Assisted in practical classes in the Algorithms and Data Structures course, given to 60 students. I was also involved in elaborating and grading the assignments.

## VOLUNTEERING

---

- **Elder digital inclusion course (6 months):** Taught the basics of how to use and interact with modern technologies, such as smartphones, cloud storage, and productivity software (Microsoft Word, Excel, etc.).