

Samuel Baptista

Ribeirão Preto, São Paulo, Brazil



samuelbaptista27@gmail.com



(19)99987-2145



[linkedin.com/in/samuel-baptista](https://www.linkedin.com/in/samuel-baptista)

Summary

Professional with 7+ years of experience in Data Science and Artificial Intelligence with two specializations in the area in addition to several courses that add up to more than 1500 hours of study.

I've worked in partnership with reference research centers in the Agro sector such as Embrapa and CPQD, contributing to the development of artificial intelligence systems that reduced costs and improved the productivity of rural producers.

I was approved among the best on LinkedIn in Machine Learning, Python, and Excel tests.

My technical expertise covers advanced concepts in the areas of BI and Analytics, Machine Learning, Data Visualization, Quantitative Analysis, Deep Learning Architectures, Computer Vision, Natural Language Processing, Cognitive Systems, and others.

I also have a solid foundation in the use of the main tools on the market for statistical analysis, database manipulation, distributed storage and processing, cloud computing, among others such as Airflow, Docker, R, SQL, Spark, Hadoop, Tableau, PowerBI, AWS, Azure, Python and its main libraries: Numpy, Scipy, Pandas, Scikit-Learn, Tensorflow, Keras, Matplotlib and Seaborn.

Portfolio:

github.com/samuelbaptista

Datacamp (80+ courses and 6 projects):

datacamp.com/profile/samuelbaptista

Experience



Data Scientist - Generative AI

iFood

Mar 2023 - Present (9 months)

- Expertise in Generative AI: Proficient in developing and implementing generative AI models to solve complex problems.
- Innovation Strategy: Experience in driving innovation within teams, including the application of cutting-edge AI technologies.
- Application Development: Skilled in developing applications with embedded Language Models.
- Team Collaboration: Proven ability to work effectively in diverse, cross-functional teams.
- Problem Solving: Strong problem-solving skills, with a focus on innovative solutions and continuous improvement.



Data Scientist - Catalog

iFood

Jun 2022 - Mar 2023 (10 months)

- Lead initiatives of data enrichment for catalog items with a focus on preparing a product sheet using state-of-the-art techniques in natural language processing (NLP) and computer vision.
- Software development using Python language (Sklearn, Pandas, Numpy, PyTorch, Keras, Tensorflow, Streamlit, Matplotlib)
- Data pipeline development using Databricks (SQL, Pyspark)
- Pipeline orchestration using Airflow
- Deployment and maintenance of models in production using AWS resources (s3, SageMaker, EC2, ECR)
- Maintenance of repositories and CI/CD tools (Git/GitLab)



Data Scientist

Banco Modal

Dec 2021 - May 2022 (6 months)

- Own new and existing initiatives around Machine Learning and Analytics.
- Lead the life cycle of Machine Learning including data transformation, exploratory data analysis, feature engineering, model trial pipelines, measuring performance, retraining, and deploying/ productizing models
- Act as a thought leader across multiple business functions to formulate DS-centric opportunities that lead to defined results
- Manage and mentor a growing team of data scientists



Data Engineer

Keyrus

Nov 2020 - Nov 2021 (1 year 1 month)

- Development of ELT processes using Databricks.
- Development in Python for communication through APIs and access to SAP system tables.
- Construction of workflows for automation of ELT flows with Airflow.
- Support to the business area for mapping the relationship rules between data.
- Database and Workflow modeling for flow optimization in a distributed processing environment.
- Data lake management using delta tables in AWS cloud environment.
- Setup and construction of data pipelines on Microsoft Azure environment.



Machine Learning Engineer

BemAgro

Mar 2020 - Nov 2020 (9 months)

- Reporting to the CTO.
- Software development using Python language and S.O.L.I.D.
- Main libraries used in the project: Tensorflow, Keras, Pandas, Numpy, OpenCV, Skimage, Sklearn, and Flask.
- Worked with the RD&I team to prepare agricultural reports from drone and satellite images (plant counting, weed detection, mapping of lines, and planting failures).
- Implementation of deep neural network architectures and other machine learning techniques for computer vision.
- Planning and execution of activities related to data collection in the field.
- Management of collected data, following good storage and labeling practices.
- Conducting static tests for project validation.

- Construction of continuous improvement pipelines in applications (CI/CD).

Relevant Projects

- Software development, approved in production, in just nine months.
- Automation of plant counting reports with estimated yield for 3 crops (Citrus, Coffee, and Eucalyptus).
- Automation in the identification of weeds for localized application of herbicide, with the result prepared to be consumed directly by agricultural machinery API, making the process 100% automated.



Data Scientist

Skylux

Jun 2016 - May 2019 (3 years)

- Reporting to the Executive Director, and responsible for a team of 6 employees
- Predictive modeling for operational demands
- Control of scenarios, risk metrics and uncertainties, statistical significance for investment analysis and validation.
- Strategic, tax, and financial planning consultancy
- Performance in business intelligence: KPI and numbers by department
- Preparation and monitoring of annual budget
- Cost control and Pricing

Results:

- Reduced billing time from 7 to 2 business days
- Increase of 6% in the company's operational efficiency - EBTIDA

Education



USP - Universidade de São Paulo

Bacharelado, Química Forense

2008 - 2011



Uniderp

Tecnologia, Gestão Comercial

2015 - 2017

Licenses & Certifications



Data Scientist with Python - DataCamp



Data Scientist with R - DataCamp



Formação Cientista de Dados - Data Science Academy



Formação Inteligência Artificial - Data Science Academy



Quantitative Analyst with R - DataCamp



Machine Learning Scientist with Python - DataCamp



Data Engineering with Databricks - Databricks



Scalable Machine Learning with Apache Spark - Databricks



Astronomer Certification for Apache Airflow - Astronomer



EF SET English Certificate 83/100 (C2 Proficient) - EF Standard English Test (EF SET)



Microsoft Certified: Azure Data Scientist Associate - Microsoft

Issued Mar 2022 - Expires Mar 2023

Skills

Machine Learning • Python • Ciência de dados • BI • Big data • Deep Learning • Artificial Intelligence (AI) • Data Analysis • Distributed Systems • Microsoft Excel