JOÃO PAULO SCHIAVON

Software Engineer/ Data Scientist

Florianópolis, BR | **hey@joaopaulo.me** | **https://joaopaulo.me**

A dedicated and curious Software Engineer with a passion for unraveling the intricacies of data to drive impactful solutions with a robust four-year tenure in the Brazilian tech industry. My journey has been steeped in the intricacies of data science projects, encompassing the complete spectrum from data processing and collection to the development of comprehensive data models. Proficient in Python and adept at navigating the AWS environment, my expertise has been honed through hands-on experience and a dedication to technological innovation.

WORK EXPERIENCE

COMPASSO.UOL – FLORIANÓPOLIS/BRAZIL

DATA SCIENTIST | Nov 2023 - Current

- Spearheaded the analysis of customer data, extracting actionable insights that directly influenced strategic business decisions.

- Employed a clustering model to identify patterns within customer datasets.

- Leveraged data visualization techniques to present complex information, improving team comprehension and accelerating decision-making processes.

SOFTPLAN – FLORIANÓPOLIS/BRAZIL

DATA SCIENTIST | Nov 2020 - Nov 2023

- Created a powerful ETL pipeline designed to preprocess raw documents, capable of handling over 1 million documents daily, utilizing AWS Python Lambda function, SQS, and S3

- Started implementing different approaches to summarize documents using the LLM model, which will help minimize data processing and enhance data quality.

- Designed a robust pipeline capable of classifying over 1.5 million documents daily using FastAPI(Python), RabbitMQ, and Celery.

- Develop various NLP models for Brazil's judicial system to perform tasks like document classification, NER, relation extraction, etc. Assisting with complicated and lengthy processes.

- Created CI/CD pipeline with DVC for versioning dataset and models to streamline training and deployment, improving the efficiency of model development and enhancement.

- Developed an annotation platform based on a fork of Doccano that allowed users to create, share and develop ElasticSearch queries as document classifiers using percolation queries. This platform helped non-technical users to develop more than 150 document classifiers. The project has a frontend in Vue.js, a backend in Django(Python), and a PostgreSQL database.

FINCH SOLUÇÕES - BAURU/BRAZIL

DATA SCIENTIST | Mar 2020 – Nov 2020

- Streamlined document processing and model training by implementing an Airflow pipeline that schedules weekly jobs and updates input data.

- Develop an integrated NLP model for classifying clusters of complex legacy documents, enabling clients to efficiently organize and observe historical data.

- Streamlined NER tasks by building a multiprocess jobs Ray (Python), resulting in 25% faster processing time.

NINDOO - SÃO PAULO/BRAZIL

PYTHON DEVELOPER | Jun 2019 – Mar 2020

- Spearheaded the extraction of valuable data from diverse government portals by deploying advanced web crawlers.
- Leveraged Selenium and Scrapy (Python) to adeptly and efficiently extract data from a range of portals
- Orchestrated the project using RabbitMQ for streamlined coordination and Celery for the effective distribution of tasks.

LANGUAGES

Language	Proficiency
Portuguese	Native speaker
English	Advanced
Japanese	Basic

EDUCATION

UNIVERSIDADE ESTADUAL PAULISTA JÚLIO DE MESQUITA FILHO

Bachelor, Computer Science | May 2015 - Jun 2019