Alysson Guimarães

Sergipe, Brazil | alyssonalk@gmail.com | linkedin.com/in/guimaraesalysson/ | github.com/k3ybladewielder/

RESEARCH INTERESTS

My research focus is artificial intelligence (AI), specifically language models. My work explores Natural Language Processing (NLP) techniques applied to healthcare, medical and biomedical domains to improve processes, reduce cognitive load on professionals and information overload through methods such as Text Summarization and Named Entity Recognition (NER). In addition, I am interested in Commonsense Reasoning in language models.

EDUCATION

Universidade Federal de Sergipe (UFS) 2024 – Present
Master in Computer Science
Advisor: Methanias Colaço Junior.

Universidade Tiradentes (UNIT) 2019 – 2020
MBA in Business Management and Competitive Intelligence
GPA: 3.64/4

Faculdade São Luís de França (FSLF) 2015 – 2019
Bachelor in Business Administration
GPA: 3.36/4

PUBLICATION AND SCIENTIFIC WORK

1. **A. Guimarães***, M. Colaço Junior. *A Systematic Mapping of Competitive Intelligence Applications based in Natural Language Processing*, **(In progress)**
2. **A. Guimarães***, M. Colaço Junior. *A Systematic Mapping on Text Summarization Methods Applied in Health Domain*, **(In progress)**
3. **A. Guimarães***, M. Colaço Junior. *Extractive Generic Multi-document Summaries Based in Topics Using Memetic Algorithms*, **(In progress)**

EXPERIENCE

GAVB Consulting – Canoas, RS, Brazil Aug 2021 – Present (Remote)
Data Scientist

- **Problem:** Low accuracy in the detection of quantities (units) in customer orders. **Solution:** Based on the order description (in weight or footage) compared to the product description, decision rules are applied to identify quantities (Ex. 150kg of product A -> 2 units of A of 100kg) with a hybrid approach using symbolic AI and LLM. **Outcome:** I improved the accuracy of the four main categories from 16% to 90%, 36% to 96%, 17% to 90%, 85% to 98%.
- **Problem:** Low accuracy in information extraction relative to products and quantities in PDF files. **Solution:** Using Pypdf and LLMs like GPT-4o mini, I designed, analyzed and implemented a framework to extract the necessary content page by page. **Outcome:** I improved the performance from 12.56% to 66.66% in correct descriptions and from 44% to 60.38% in product quantities for an e-commerce.
- Benchmarking between LLMS for generating budgets.
- Statistical analyses to assess the impact of implemented AI-based systems, verifying the effects correlated to use and causal impact
- **Problem:** Analyzing the content of consumer interviews was a manual, time-consuming, highly subjective process that required +10 professionals, taking approximately one month to complete. **Solution:** I Developed of a product for sentiment analysis and topic modeling in interviewees' statements, detecting the main problems/demands reported. **Outcome:** Just considering the number of hours of the manual process with +10 people, when compared to the execution time of the framework, I reduced the analysis time by approximately 60%.
- Work on a forecasting project, assisting in the development of the algorithm and generation of results, in which they supported the company's strategic planning for the next 10 years.

- Development of a Proof of Concept (POC) for a sales assistant based on LLMs. The assistant interacted with the consumer, indicating products based on preferences and explaining how the product was suitable for the customer, linking to the product in ecommerce.
- **Problem:** In the civil construction sector, serving all customers (leads) is a time-consuming process, aggravated by the lack of resources such as available real estate agents, limited working hours and high volume of leads. **Solution:** I developed a property purchase propensity detection algorithm, to prioritize leads that are more likely to purchase, while those with a low propensity would return to the sales funnel and receive more information about the property of interest. **Outcome:** Potential for revenue growth, through customer prioritization, given limited resources.
- Applied machine learning for regression, classification, and clustering using algorithms like linear regression (statsmodels), CatBoost, and k-modes.

TECHNOLOGIES

Languages: Portuguese (Native), English (B2)
Programming Languages: Python, SQL
Machine Learning Tools: Scikit-learn, CatBoost, NumPy, Pandas, Seaborn, Plotly, PySpark, Databricks
Technologies and Skills: Full Data Science Pipeline, Power BI, Google Data Studio, Regression, Classification, Time Series Analysis, Hypothesis Testing, Excel, Git, Airflow
Cloud: Knowledge in Databricks, AWS, GCP, Azure (Development to Deployment)