NLP – Studing

Sentimental Analysis with Supervisioned Learning

A diagram of a function

Description automatically generated

Text to a Vector

You need to create a vocabulary with all words that is present on your texts.

Sparse Representation –

A screenshot of a computer

Description automatically generated

A white background with blue text

Description automatically generated

**Keras**

A diagram of a diagram

Description automatically generated with medium confidence

A diagram of a computer process

Description automatically generated

A screen shot of a computer

Description automatically generated

A graph of a line with arrows

Description automatically generated with medium confidence

Oh nice to meet you.

STAR – **Situation**, **Task**, **Action**, **Results**

***Questions***

**Tell me about yourself.**

Thank you for the opportunity to be interviewed for the position today with your company, which I am very excited about.

As you know, my name is Marcio, I am a Senior Data Scientist with over 6 years of experience, I build up a diverse range of skills applied to Data Science and Machine Learning field and I believe I’m the best match for this position.

I specialize in building data pipelines and applying machine learning and deep learning models for tasks like NLP, classification, forecasting, clustering and I consider myself a Python Specialist. I have hands-on experience with Python, PySpark, and AWS services such as EMR and SageMaker. I've worked on projects involving frameworks like Hugging Face, spaCy, and OpenAI.

Tell me about some Project

Situation

At my current job, our manager noticed that banks in Brazil often struggled to efficiently interpret and comply with government regulations due to the complexity and volume of the documentation. This created bottlenecks for compliance teams and increased the risk of errors. I was tasked with developing a solution to streamline the process of reading, classifying, and querying these regulatory documents, ensuring that teams could quickly access accurate information without manually reviewing the lengthy PDFs.

To address this, I:

Used word embeddings and a pre-trained GPT-4 model to build a chatbot capable of understanding complex regulatory language.

Applied prompt engineering to tailor GPT-4’s responses for precision when answering questions related to regulations.

Implemented a semantic analysis model from Hugging Face to classify documents into appropriate categories.

Performed topic modeling with LDA to identify the main themes of the documents.

Enhanced the model's contextual understanding by integrating synonyms and related terms to improve its performance on nuanced queries.

The solution significantly reduced the time compliance teams spent searching and interpreting regulatory documents. The chatbot achieved an75% accuracy rate in document classification and provided reliable answers to user queries, improving efficiency and reducing operational risks for the organization.

**BLEU Score:** To measure the similarity between the chatbot's responses and human-annotated answers, focusing on semantic relevance.

"At Capco, I worked on a natural language processing project where I developed an application to read and analyze PDF documents. The application extracted content from PDFs, broke down the text into chunks or vectors, and used a semantic search class that I implemented to find relevant information within the content.

Our application then provided not only the relevant context but also synonyms and topics related to the question, further enhancing the accuracy of responses. This combination of data was sent to GPT via an API, and the returned responses were significantly more accurate thanks to the rich context provided.

This project gave me hands-on experience with prompt engineering, NLP integration, and embedding techniques, along with familiarity using models like GPT via API. With this experience, I’m confident that I can add value to your AI Engineering team.”

"At Itaú, I worked on a project for classifying tables based on user comments. Our goal was to automatically identify which business group each table belonged to. To achieve this, we conducted a prior data curation process and then built a supervised model using TF-IDF to transform the comments into vectors, capturing relevant information for classification. This approach enabled efficient classification of the tables, making it easier for different business sectors to access specific data.

**Why do you want to work for us?**

I have studied the job and my skills, qualities, and strenghs are an excellent match for the position. I have researched your company and am atrracted to your values, positive workplace culture, and how you offer an inclusive environment that gives equal opportunity to everyone.

**What are your greatest strengths?**

My strengths include being a supportive collaborator, taking ownership of difficult problems and setting a good example for others to follow. I’m very good with clients and custumers because I understand for your businnes to be successful I need to provide outstanding service. Also I’m very fast learner, so if you give me a task, I will learn it quickly and I always will be the person who will volunteer.

**Why should we hire you?**

You should hire because I am a quick learner, require minimal supervision in my work, and will take responsibility for my professional development, wich will give you a great return from my salary . I’m a positive person, I have goals outside of work, which means that I need to perform to a high standard when I’m in my job in order to achieve those goals.

**What is your biggest weakness?**

**Where do you see yourself in 5 years?**