André Provensi

1993-04-27
São José - SC - Brazil
andreprovensi@gmail.com | +55 48 9986-77118

LinkedIn: https://www.linkedin.com/in/andreprovensi/
GitHub: https://github.com/andreprovensi

Summary

Experienced data scientist with a proven track record of leveraging Python, PySpark, Pandas, Numpy, Scikit-learn, SQL, AWS, and PowerBI to extract valuable insights and drive data-driven decision-making. I have successfully applied data science, AI, and machine learning to tackle business challenges such as fraud detection, customer segmentation, and understanding customer behavior and product relationships.

I have collaborated closely with cross-functional teams in fraud, marketing, and product, developing my skills in ETL processes and data integration. My work involved wrangling diverse datasets, ranging in size and complexity, to uncover actionable insights and develop robust machine learning models.

I also have good interdisciplinary communication skills, demonstrated through my ability to explain technical concepts and present insights to diverse audiences with varying backgrounds.

Experience

Data Scientist *December 2022 - Current moment*
DM Card

DATA SCIENTIST RESPONSIBLE FOR:

- **Evaluation of alternative data to enhance model performance and generate new insights**
  A supplier company gave the company where I work a sample of data collected from customers' smartphones when they access our app and the company wanted to evaluate the potential of this data to generate new insights about our customers and to improve the performance of some machine learning models already in production.

First I have developed an ETL process in PySpark to clean the data. After that I have crossed the new data with our internal database and explored the potential of the alternative data to predict customer churn. I have developed a machine learning model that had 70% accuracy predicting churn. For this model I have developed a feature engineering process with the customers' app that showed great potential to improve the performance of the current credit default model. I have also conducted A/B tests based on the customers' apps.
My analyses were able to show the potential of the alternative data and the company signed a contract to regularly obtain the data collected from the usage of the app.
Skills: AWS glue, AWS athena, SQL, Python and PySpark.

- **Anomaly detection for internal fraud:**
  The fraud department wanted to monitor and identify suspicious behavior of people using the company's internal system and there was no tool for that purpose.
  Since this kind of fraud had never been evaluated at the company, the fraud team and I worked together to develop a solution. After deciding what kind of data to use and features with capability to detect fraudulent behavior and misuse of the system, I developed an anomaly detection machine learning system based on the logs of the company's system that generates fraud alerts. The logs were collected from the AWS cloud by SQL queries and two probabilities scores were developed: one using a Gaussian distribution and other using a Poisson distribution.
  The machine learning system detected more than 50 cases in two months after deployment.

- **Product team data scientist**
  The company decided to place data scientists in business areas in order to use data science to improve business results using machine learning and performing studies to turn data and insights into actions and I was chosen to be the data scientist for the product team.
  I have performed thorough studies and data analysis regarding customer behavior and interaction with the company's products. Using unsupervised learning I have identified customers profiles and formulated hypotheses. Together with Product Owners, we are planning actions with different squads based on the results found by my studies.
  Skills: AWS glue, AWS athena, Python and PySpark.

- **Business Intelligence for fraud rules and alerts:**
  I worked together with the fraud department to automate a time-consuming manual process for generating reports about efficiency of fraud rules and amount of fraud alerts for the company's financial products. After discussing and understanding their needs, I have developed a PowerBI dashboard that automatically collects the necessary data using SQL

queries on AWS Athena and displays business metrics ready for analysis and decision making for the fraud department. The automation of the dashboard allowed the fraud department to stop spending time to collect data and build reports, enabling them to invest their time in analyzing the efficiency of rules and to assess in a more agile way which rules needed to be updated or replaced.

Python Software Developer *August 2022 - October 2022*
Freelance

SOFTWARE DEVELOPER RESPONSIBLE FOR:

- **Development of software for measurement of mouth lesions surface area**
  Development of a software in Python for academic research at the Federal University of Santa Catarina in the field of dentistry. The software accurately calculates the surface area of mouth lesions from user-uploaded photographs, correcting for perspective distortion caused by the camera angle. Users can select known dimensions within the image and use one of three methods (freedraw, polygon, or spline) to define the lesion border. The software then calculates the surface area of the selected region. You can check the software on https://github.com/andreprovensi/UI-area-image

Systems Engineer / Systems Technician *September 2021 - December 2022*
Akaer Engineering

SYSTEMS ENGINEER RESPONSIBLE FOR:

- **Military Training Drone Cooling, Electrical and Recovery Systems**
  I was directly responsible for the cooling system, but I worked with other systems engineers from my team to solve problems of other systems.
  During the development of a military training drone, a series of requirements provided by the customer had to be fulfilled. Solutions had to be proposed by us and we also had to assess whether or not these solutions fulfill the requirements.
  Regarding the cooling system, I developed physical and mathematical models in Python to evaluate the efficiency of thermal protection solutions of temperature sensitive components of the drone proposed by our team. My models described with precision the behavior of the drone components and were used to choose the best thermal protection solution.
  Regarding the electrical system, I have performed random vibration simulations using finite element analysis with FEMAP to simulate the behavior of an electric component during an experiment. The results of my simulation showed that the component was able to be tested for vibration.

Regarding the Recovery System, I have performed statistical analysis to determine the rupture tension of a series of possible tissues that would be used to manufacture an airbag. The statistical analysis was able to select the most resistant tissue with statistical significance for the airbag production.

Skills

- Python
- SQL
- Machine Learning
- Data Mining
- Analytical Problem Solving
- PowerBI
- Data Analysis
- Statistic

Languages

- English - Fluent
- Portuguese - Native
- German - Intermediary
- Italian - Basic

Certifications

- Data Science Professional Certificate - Issued by IBM on May 2023
- Machine Learning Specialization - Issued by Stanford University, DeepLearning.AI on October 2023