Candidatul ideal

eMAG is looking for a highly motivated colleague to join our AntiSpiders team as Data Scientist.

Our team analyses web traffic, identifies non-human agents (crawlers, search engines) and develops pipelines for filtering data used by the Business Intelligence department.

We have built our applications with Spark, Kafka, BigQuery, and Airflow. They are mostly written in Python and deployed using Docker on a hybrid infrastructure (own machines and Google Cloud)

You will analyze large-scale datasets and build machine learning models to enhance spider detection. These models will be integrated in our production pipelines.

Descrierea jobului

Responsibilities

Discover, interpret and document unique insights in large-scale distributed datasets through exploratory analysis and the application of advanced analytical methodologies;
Create statistical and machine learning models in multiple technologies using best-in-class data science approaches;
Processing, cleansing, and verifying the integrity of data used for analysis;
Develop dashboards and reporting workflows for internal clients;
Write production-level code for training and inference: data validation, logging, exception handling, and orchestration.

Requirements

A degree in a related discipline (Mathematics, Statistics, Data Science, Computer Science);
1 year of related work and/or research experience in quantitative roles;
Knowledge of at least one open-source scientific or statistical programming language such as R or Python;
Fluency in SQL and good knowledge of relational databases;

Experience with data visualization tools.

The successfull candidate will:

Be passionate about asking and answering questions in large, distributed datasets;
Be a clear and confident communicator with the ability to translate technical concepts into a concise business focused messages in both technical and management presentations;
Have knowledge in the following data science domains: Anomaly Detection, Time Series Analysis, Unsupervised Learning;
Have the ability to collaborate with an interdisciplinary team to solve problems;
Have a good understanding of the design and architecture of big data applications;
Possess a solid understanding of the value of data and how technology enables companies to compete better in the market place.