Senior Data Engineer

Foreword

This is just a test to evaluate your average skills and experience. There isn't a unique way to succeed in it. Different candidates can succeed in many different ways. Just try to do your best while enjoying it. We are a tech company that deeply believes in data driven processes and a data driven mindset is what we look for.

Theory assignment

Answer the following questions by giving us some details and arguments to better understand your point of view (10 lines max).
1. What is the most efficient way for a Data Engineer to work with Data Scientists, Data Analysts and Software Engineers?
2. Which was the most game changing learning you ever had working on massive data pipelines? Why was it so crucial under a technical and business point of view?

Technical assignments

Context

A team is going to launch a large scale mobile app, where content recommendation is a key part of the product.
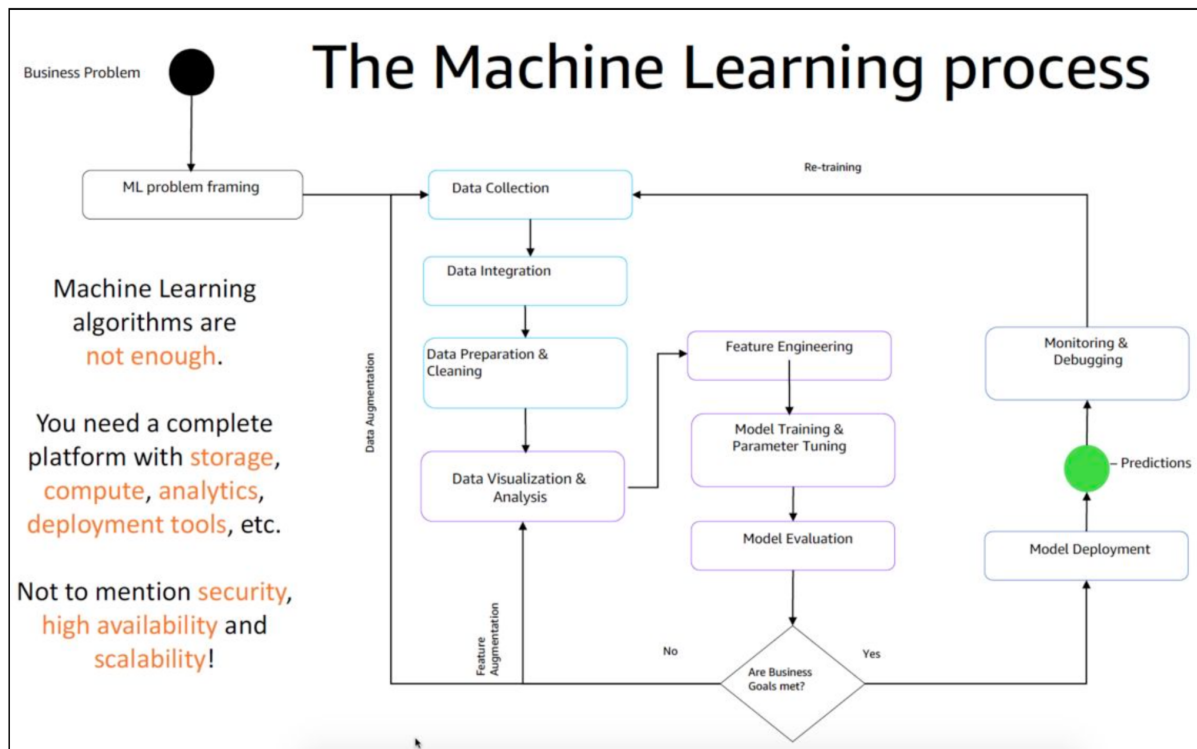The app will generate billions of daily events (10TB), with millions of DAU.
The backend of the app uses AWS, Node.js and some serverless functions.

As a Data Engineer, you will work closely with Data Scientists to develop new ways to acquire data, build processes to create data sets, setup machine learning infrastructure, etc.

Content

Here is a slide that sums up a Machine Learning process, and outlines the importance of having not only algorithms, but a complete infrastructure to support it.

So far, the team only uses a third-party tool (Google Analytics) and is looking to build an in-house data pipeline to work on Machine Learning and execute the above process.

Assignment

1. Design the infrastructure to support all steps of this "Machine Learning process". This design document includes a list of stacks that need to be created, the technologies/services used, pros/cons, diagrams, …
2. Prioritize the work in a high-level roadmap, with timing, dependencies, …

Deadline

We consider 5 hours to be enough to pass the test. But this is just a suggested deadline. It's ok if you need more time to achieve a better result. In this case just write us an email where you explain the reasons for the delay.

Evaluation method

We are going to evaluate your project under different criteria:
- creativity
- technical coherence and relevance

- business coherence and relevance
- ability to communicate your findings in a structured and efficient way, to your different stakeholders
- ability to organize & prioritize your work

Provide a well written test and consider your results will be shared internally with all the stakeholders of the data team. So try to be as clear as possible. If you encounter any problem please feel free to contact us by email.

Hand in

Once you finish your test, please send it to stefaniec@voodoo.io.