

Tema:

Arquitetura de um sistema de recomendação baseado em uma rede neural

Contexto

Atualmente se observa uma dificuldade durante o ciclo de vida de desenvolvimento de projetos em alinhar as necessidades de um negócio e suas demandas da área de ciência de dados.

Muitas vezes ocorre uma desconexão entre os modelos de inteligência artificial gerados pelos cientistas de dados e sua implementação prática em aplicações reais. Quem sofre com isso são os usuários desses sistemas que não tem sua experiência diretamente afetada pelo que foi desenvolvido.

Objetivo

Desenvolver uma arquitetura capaz de integrar um sistema de recomendação baseado no aprendizado de máquina e redes neurais com uma aplicação web dedicada a relacionamentos.

Os usuários que interagem com a aplicação devem ser capazes de avaliar os outros perfis de clientes a partir das imagens e assim, após o treinamento do modelo, receber uma recomendação do sistema.

O modelo treinado terá um dashboard de avaliação disponível para um usuário administrador especialista fazer a análise dos resultados de performance e possivelmente fazer ajustes ou aprovar para uso.

Metodologia

Para a execução do projeto foi utilizado o framework de *design science*, além das metodologias estudadas no laboratório de engenharia de software para o desenvolvimento de produtos.

Entre elas está a definição das cinco principais visões para a execução de um projeto, das quais apenas as 3 principais foram utilizadas: visão da empresa, visão computacional e visão da infraestrutura.

Tendo em vista as principais necessidades do usuário do sistema e como impactar sua experiência da forma mais significativa através de um algoritmo de recomendação inteligente se estabeleceu a arquitetura do sistema.

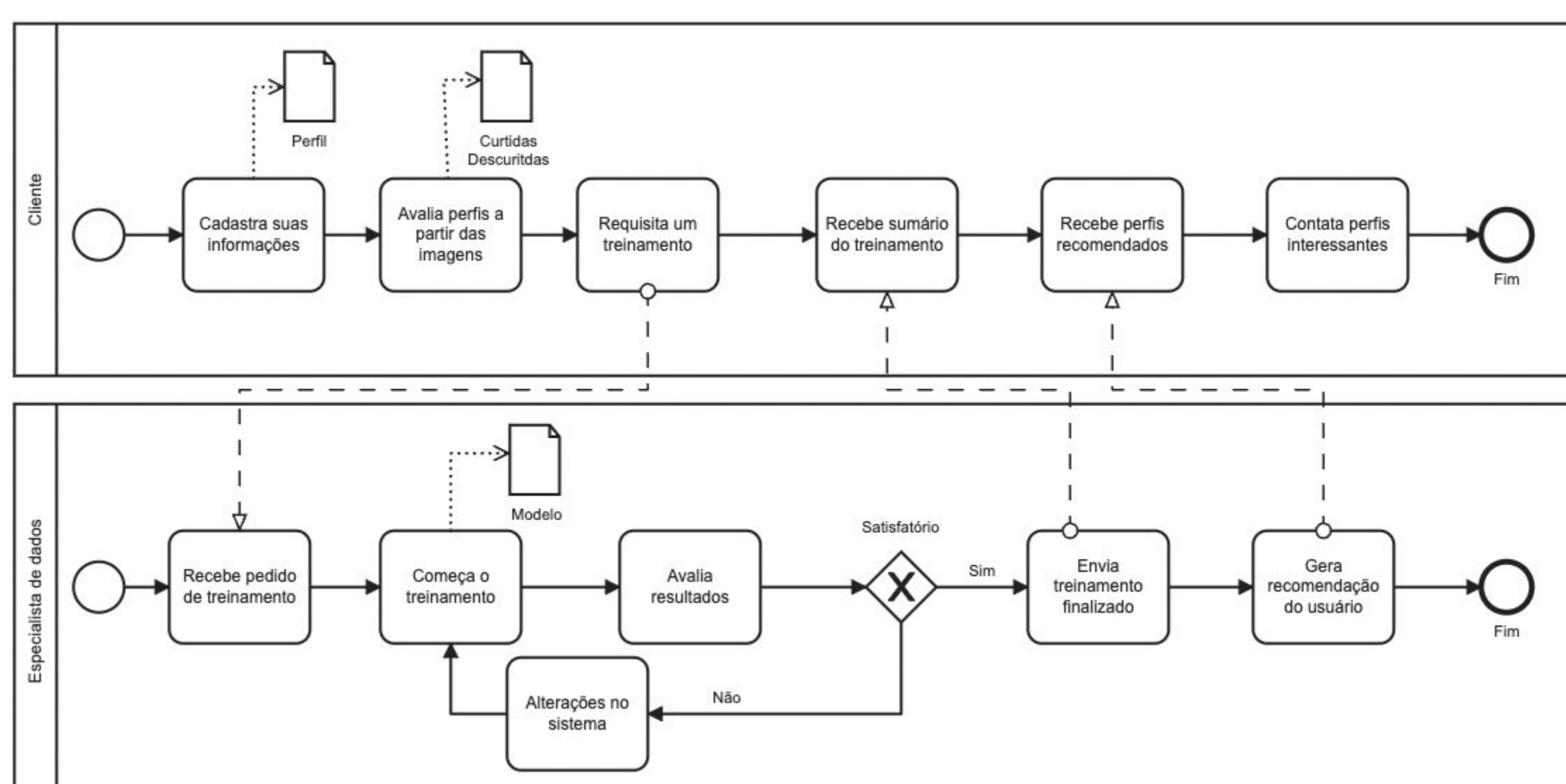


Figura 1: Modelagem BPMN do fluxo principal

Fonte: Autor

Integrantes: - Eduardo Castilho de Almeida Prado

Professor(a) Orientador(a): - Prof. Dr. Jorge Luís Risco Becerra

Arquitetura

A arquitetura do projeto é composta por 3 camadas. O *front-end* foi desenvolvido em **React** web utilizando *Javascript*, e o site está sendo hospedado no Netlify.

O *back-end* foi desenvolvido utilizando **Python**, sendo o framework **Flask** utilizado para a parte do servidor web e um **Jupyter notebook** utilizando **TensorFlow** para a parte de aprendizado de máquina. O servidor **Flask** está hospedado no Heroku, que contém um banco em **Redis** para autenticação dos usuários.

O banco de dados utilizado para os dados dos usuários foi o **PostgreSQL** que está hospedado em uma instância do **Amazon RDS**. Já os bancos de dados que contém as imagens, modelo de machine learning e dados de performance são instâncias do **S3**.

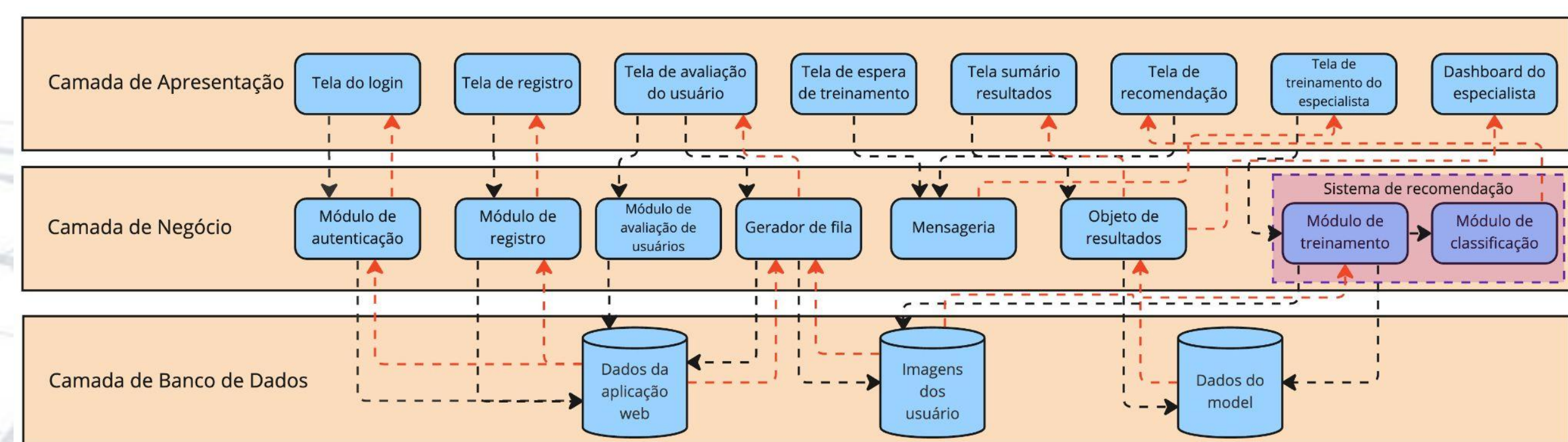


Figura 2: Arquitetura de camadas do sistema

Fonte: Autor

Implementação

A rede neural foi desenvolvida utilizando a técnica de *transfer learning*, tendo como base um algoritmo com alta acurácia já implementado pela Google para classificação de imagens, o **Inception V3**.

O modelo é modificado e calibrado para classificar as imagens binariamente utilizando o banco de dados montado pelo usuário enquanto ele avaliava os perfis como "gosto" ou "não gosto".

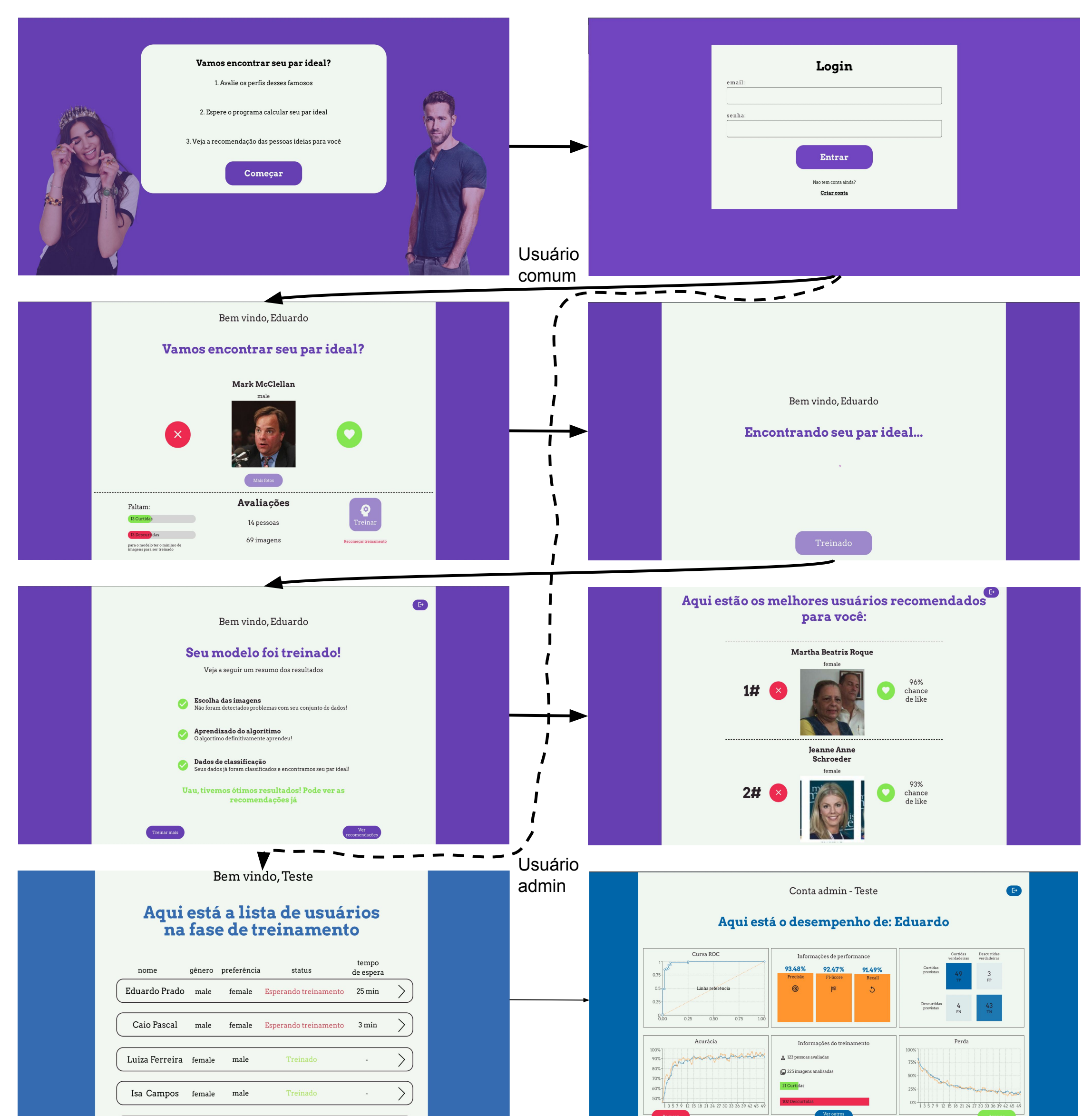


Figura 3: Tela da aplicação com fluxo do usuário e admin

Fonte: Autor