Data Scientist Co-Op Case Study

We're excited to learn more about how you approach data problems! This case study will give you the opportunity to showcase your skills. You'll have 48 hours from the time you receive this case study to submit a completed response. During your interviews, be prepared to present your findings to us with a slideshow.

Context

Cohere Pets is on a mission to simplify veterinary care for pet owners and veterinarians, aiming to ensure pets receive necessary procedures without delay, thereby preventing major health risks. When a pet requires a procedure, veterinarians must obtain prior authorization ("auth") from the pet owner's insurance company before proceeding. This ensures that the procedure is covered by insurance and is clinically appropriate for the pet. Typically, the insurance uses medical claims history to determine whether an auth is approvable or not.

Currently, to approve or deny auths, Cohere Pets uses rules, which are essentially "if/else" statements that auto-approve auths if the veterinarian correctly answers a set of questions that determine clinical appropriateness. To streamline this process, Cohere Pets' Data Science team wants to also develop machine learning models that automatically approve some prior authorizations at the time of submission. Cohere Pets generates revenue by managing these prior authorizations on behalf of insurance companies.

Workflow

1. A pet owner takes their pet to a veterinarian.
2. The veterinarian determines that the pet needs a procedure.
3. The veterinarian's office staff uses Cohere Pets' portal to submit a prior-authorization request, uploads supporting documentation and answers clinical questions.
4. Cohere Pets' backend rule engine automatically approves some authorizations based on the answers to the clinical questions. If an authorization is not auto-approved, it goes through human review.

5. Once approved, the veterinarian performs the procedure and submits the claim to the insurance company for reimbursement.

Datasets [link](#)

- **Prior Authorization Data**: Historical data on prior authorization requests, including decision outcomes.
- **Claims Data**: Historical data on medical procedures received by pets, sourced from insurance companies.

Exercise

1. **Identify Business Problems**: What potential business challenges could Cohere Pets' Data Science team address using this data?
2. **Design a Predictive Model**: Develop a predictive analysis project to auto-approve prior authorizations. You should use the claims history and any other useful features you find to predict prior authorization approval.
3. **Data Processing**: Preprocess provided datasets using Python. You can join the prior authorization data and claims data using the `pet_id` and `Pet Id` columns.
4. **Exploratory Data Analysis**: Perform exploratory data analysis and create visualizations that provide insights into the data.
5. **Model Building**: Build a predictive model (or set of models) to auto-approve prior authorizations **that are not already auto-approved**. Use the `authstatus` column in the prior authorization data as the target variable.
6. **Presentation**: Prepare to present your project, including your modeling process and results. Create the presentation as if your audience were a moderately technical stakeholder. Someone who is interested in using your model(s) in the auto-approval pipeline, but isn't an expert in machine learning.

Evaluation Criteria: We'll assess your knowledge of data analysis and visualization, feature engineering, cross-validation, business sense, and communication through this exercise. Your presentation should walk us through your thought process on how you chose to solve this problem.

Please submit both:

1. A Jupyter notebook running Python code with outputs displaying your results
2. A 20-30 min slideshow