aidoc

AI Data Analysis Exercise

Our company develops deep learning solutions to optimize radiologists' workflow and expedite patient care. This exercise will focus on the Intracranial hemorrhages (ICH) detection algorithm for CT scans – a real life-saving algorithm implemented in dozens of sites across the world!
Please watch this video for further information about ICH - **Intracranial Hemorrhage**
Note: All the data in this exercise is generated and doesn't include any real operational information.

Three different algorithms have been developed, each was trained differently on the same dataset. We have deployed the three of them in two different hospitals in order to examine their performances and decide on the best way to use them.

In the hospital, a radiologist is examining (independently) the scans according to the scans queue (FIFO prioritization). We can interfere with the order of the scans in acute situations (by prioritizing scans with acute findings by the Aidoc algorithm). Furthermore, the radiologist may use our prediction (final result or the marked findings on the scan) to decide on difficult findings. When he gets to a final decision, he can conclude and sign the report.

Your task:
You were assigned to a combined team containing product, AI, and data members. The team will need to measure the effectiveness of the algorithms and offer further insights to the product and operations departments.

Product questions (max 30 min):

1. We would like you to conduct and present an initial data exploration. For example:
    a. Patients' distribution between departments (with different aspects, gender, age, etc.).
    b. Treatment duration (From scan time to the radiologist sign time).
    c. Any further exploration that may help you "feel" the data.

AI & Data questions:

1. Examine the performances of each algorithm. We would like to hear your insights about them. Tip: you're welcome to explore the common metrics for AI in this link (Specificity, Sensitivity, PPV are must-haves).
2. Compare the performances and results of the various algorithms.
3. How would you suggest improving each algorithm? Please feel free to analyze the data from different angles in order to get interesting and impactful insights.
4. You may create a document for planning future analysis - if you had more time which aspects would you analyze, what further data you may want to look through, further processes you will implement to make your analysis more valid, or other steps you would do and you didn't have the time.

Data source: excel file, attached to the mail.

Some notes for the data naming:

1. Accession - unique numbering of our data. Data that is not gathered by our CT algorithm will not have it.
2. Site - hospital's name.
3. patient_class- "ED" means it is an emergency department patient. "IN" means inpatient - regular hospitalization.
4. Gender - patient's gender.
5. Age - patient's age.
6. Scan_timestamp - time of the actual scan (this field is saved automatically by the scanner).
7. radiologist_answer - the research's ground truth. The radiologist interprets the scan by his own list of priorities.
8. report_sign_time - the time in which the radiologist finished independently interpreting the scan.
9. Algos_start_run - each scan is uploaded to our server. Once it reaches the server, all algorithms (1-3) analyze it in parallel. This field mentions the time they started analyzing the scan.
1. algoX_answer - positive (P) or negative(N) according to the findings in the scan.
2. algoX_finish_run - algorithm's end of run time.

Please analyze the data using **Python**. Try to build it as modular as you can.
The output should be saved as:

1. **Code**: A Notebook or PythonL script containing the full analysis.
2. **Slides**: A presentation summarizing the analysis task, highlighting key insights tailored for the relevant stakeholder.

Please send these to your interviewer in advance.

If you need any clarifications or have more questions - feel free to reach out.
Good luck!

AI Data Development Group,
Aidoc Medical