

Overviews of the workshop

Workshop 0: Preparatory Materials

The preparatory workshop provides crucial background information about the inaccuracies of pulse oximeters for populations with darker skin tones. It explains how these inaccuracies can lead to missed cases of hypoxemia, resulting in less treatment and higher mortality rates. The workshop also provides an overview of the datathon's schedule and recommends literature for the participants to review. Additionally, the participants are prepared for the data analysis.

Workshop 1: EDA & Study Design

In this workshop, the teams are expected to define the inclusion criteria to build a working cohort and explore the ground truths present in the dataset. The deliverables include a flow chart detailing the inclusion and exclusion criteria and the definition of the machine learning task to be addressed. The workshop also discusses potential pitfalls such as sampling and representation bias.

Workshop 2: Clinical Variables Selection & Feature Engineering

This workshop emphasizes the importance of data preprocessing in the data science workflow. Participants are taught data cleaning, normalization, transformation, and reduction techniques, and they apply these techniques to real-world datasets. The workshop also highlights how to identify and mitigate biases that can be introduced during data preprocessing.

Workshop 3: Let's get our model!

This workshop involves splitting the data into training and test sets, defining performance metrics for model evaluation, and developing a machine learning model. The models could predict arterial oxygen saturation values, the gap between arterial and peripheral oxygen saturation, or detect instances of hidden hypoxemia. The workshop also covers grid-search parameter tuning and feature importance assessment.

Workshop 4: Try to Tackle the Biases and Re-Model

In the final workshop, participants are encouraged to develop their own ideas and solutions to tackle biases and remodel the datasets based on the knowledge and skills acquired from the previous workshops. It promotes creativity and self-guided exploration and encourages participants to address various problems, such as demographic biases, missing patients from the database, developing a good fairness metric, and validating and expanding the proposed models.

ワークショップの概要

ワークショップ0：準備資料

この準備ワークショップでは、肌の色が濃い人々に対するパルスオキシメーターの不正確さに関する重要な背景情報を提供する。このような不正確さが低酸素血症の症例見逃しにつながり、治療の遅れや死亡率の上昇につながることを説明する。また、このワークショップでは、データソンのスケジュールの概要と、参加者が検討すべき文献を推奨する。さらに、参加者はデータ解析の準備を行う。

ワークショップ 1：EDA と研究デザイン

このワークショップでは、作業コホートを構築するための組入基準を定義し、データセットに存在する真実の根拠を探る。成果物には、組み入れ基準と除外基準の詳細を示すフローチャートと、取り組むべき機械学習タスクの定義が含まれる。ワークショップでは、サンプリングや表現バイアスなどの潜在的な落とし穴についても議論する。

コメントの追加 [1]: 組入に統一したい

コメントの追加 [2R1]: 了解です。

削除: 包含

ワークショップ 2：臨床変数の選択と特徴量エンジニアリング

このワークショップでは、データサイエンスのワークフローにおけるデータの前処理の重要性を強調する。参加者はデータのクリーニング、正規化、変換、削減テクニックを学び、これらのテクニックを実際のデータセットに適用する。また、このワークショップでは、データの前処理中に発生する可能性のあるバイアスを特定し、軽減する方法にも焦点を当てる。

コメントの追加 [3]: 「特徴量」で良いかもです

コメントの追加 [4R3]: ありがとうございます、そうですね。

削除: フィーチャー

ワークショップ 3：モデルを作成しよう

このワークショップでは、データをトレーニングセットとテストセットに分割し、モデル評価のパフォーマンス指標を定義し、機械学習モデルを開発します。モデルは、動脈酸素飽和度値、動脈酸素飽和度と末梢酸素飽和度のギャップを予測したり、潜在性低酸素血症の事象を検出したりします。このワークショップでは、グリッドサーチによるハイパーパラメータの最適化及び特徴量の重要性評価についても取り上げる。

コメントの追加 [5]: 潜在性

コメントの追加 [6R5]: ありがとうございます、そのとおりです。

削除: 隠れ

コメントの追加 [7]: ?

コメントの追加 [8R7]: 失礼しました、"インスタンス"とは、特定の事象や状態の具体的な発生や例を指すので、「潜在性低酸素血症の事象を検出したりします。」ではいかがでしょうか。

コメントの追加 [9R7]: ありがとうございます！良いかと思います！

削除: インスタンス

コメントの追加 [10]: ハイパーパラメータ

コメントの追加 [11R10]: グリッドサーチによるハイパーパラメータの最適化及び、ではいかがでしょうか。

削除: チューニングと

コメントの追加 [12]: "demographic bias"は「患者背景の偏り」が良いですね

削除: 人口統計学的な

ワークショップ 4：バイアスに対処し、再モデル化を試みる

最後のワークショップでは、参加者はこれまでのワークショップで習得した知識とスキルに基づいて、バイアスに取り組み、データセットを再モデル化するための独自のアイデアやソリューションを開発することが奨励される。このワークショップでは、創造性と自主的な探求を促進し、患者背景の偏り、データベースからの患者の欠落、優れた公平性の指標の開発、提案されたモデルの検証と拡張など、さまざまな問題に取り組むよう参加者に促す。

Reference: the dataset overviews

Patient Information

- subject_id: int64
- stay_id: int64
- hadm_id: int64
- gender: object
- sex_female: int64
- anchor_age: int64
- race: object
- race_group: object
- language: object
- insurance: object
- weight: float64

- height: float64
- BMI: float64
- anchor_year_group: object

Hospital & ICU Stay Information

- first_hosp_stay: bool
- first_icu_stay: bool
- icustay_seq: int64
- admittance: object
- disctime: object
- icu_intime: object
- icu_outtime: object
- los_hospital: int64
- los_icu: float64

Clinical Measurements

1. Patient Status Indicators:

- SaO2_timestamp: object
- ventilation_status: object
- mortality_in: int64
- invasive_vent: int64

2. Blood Oxygen Measures:

- SaO2: float64
- delta_SpO2: int64
- SpO2: int64
- hidden_hypoxemia: int64

3. General Health Indices:

- CCI: int64
- SOFA_admission: int64
- delta_vent_start: float64
- delta_rrt: float64
- rrt: int64
- delta_vp_start: float64
- norepinephrine_equivalent_dose: float64

4. SOFA (Sequential Organ Failure Assessment) Score Components:

- delta_sofa_coag: float64
- sofa_coag: float64
- delta_sofa_liver: float64
- sofa_liver: float64
- delta_sofa_cv: int64
- sofa_cv: int64
- delta_sofa_cns: float64
- sofa_cns: float64
- delta_sofa_renal: float64
- sofa_renal: float64
- delta_sofa_resp: float64

- sofa_resp: float64

5. Blood Cell Measures:

- delta_hemoglobin: float64
- hemoglobin: float64
- delta_hematocrit: float64
- hematocrit: float64
- delta_mch: float64
- mch: float64
- delta_mchc: float64
- mchc: float64
- delta_mcv: float64
- mcv: float64
- delta_platelet: float64
- platelet: float64
- delta_rbc: float64
- rbc: float64
- delta_rdw: float64
- rdw: float64
- delta_wbc: float64
- wbc: float64

6. Coagulation Measures:

- delta_d_dimer: float64
- d_dimer: float64
- delta_fibrinogen: float64
- fibrinogen: float64
- delta_thrombin: float64
- thrombin: float64
- delta_inr: float64
- inr: float64
- delta_pt: float64
- pt: float64
- delta_ptt: float64
- ptt: float64

7. Liver Function Measures:

- delta_alt: float64
- alt: float64
- delta_alp: float64
- alp: float64
- delta_ast: float64
- ast: float64
- delta_bilirubin_total: float64
- bilirubin_total: float64
- delta_bilirubin_direct: float64
- bilirubin_direct: float64
- delta_bilirubin (assuming this might be