OneModel Academy

One AI

Per Column Interventions

Machine Learning

Austin Hambrick
Machine Learning Engineer

12 Jul 2024

Topics Covered

- Overview of per column interventions in machine learning
- Types of per column interventions available to use with One AI models & when they are most appropriate to utilize
- How to configure each type of per column intervention in One Model

Learning Outcomes

You will:

- Understand what per-column interventions are when & how to apply different per-column interventions ensuring each feature is optimally processed
- Configure column droppability settings and select appropriate null-filling strategies to maximize the use of available data & improve model accuracy
- Apply tailored interventions for categorical & continuous features, ensuring that each type of data is handled effectively to enhance overall model performance & interpretability

Per Column Intervention Overview

Per Column Interventions (PCIs)

- Specific preprocessing & transformations to individual features in the model dataset, vs. applying the same preprocessing to the entire dataset
  - Different columns have unique characteristics & requirements, calling for tailored preprocessing
  - Actions include cleaning, transforming, handling missing values, normalizing or scaling data, encoding categorical variables, etc.
- Improve model performance & interpretability
  - Models can better capture the underlying patterns & relationships in the data
  - Reduces the risk of overfitting
  - Better handles diverse data types & missing values

Per Column Interventions Available in One AI

Droppability

3 options

- **Is droppable:** default setting; column can be selected if predictive, or can be dropped by One AI if it violates global settings or does not improve performance
- **Not droppable:** column can’t be dropped by One AI
    - Doesn't guarantee the column will be selected, but it will always be processed & considered, even if it violates global settings
- **Always:** column will always be dropped & not tested in any way

Null-Filling

Good alternative to changing droppability or increasing the null drop threshold

Strategies:

- **Mean / Median**: the average or middle value of a continuous column are used to fill the null values; can't be used on categorical columns
- **Mode**: the highest occurrence value are used to fill the null values
- **B-fill:** null values are filled starting from the last value, working toward the first value; any null is filled with the next populated value
- **F-fill / Pad**: opposite of bfill; null values are filled starting from the first value, working towards the last value; any null is filled with the previous populated value
- **Custom:** use any custom fill value (numerical, categorical, or date) to fill all null values

Type-Specific Interventions

To customize how individual columns are processed & included in the model

- **Categorical**
  - **Exclude -** enter the values to exclude from being considered as features
  - **Select -** enter the values that the model will then force select as features
- **Continuous**
  - **Force Select -** force the model to select a variable as a feature
  - **Continuous strategy -** change the continuous strategy for the specific column to bin or scale

Configuring in One AI

OneModel Academy

Thanks for watching!

OneAI