

PROJECT SPECIFICATION

AI and EHR

Exploratory Data Analysis

CRITERIA	MEETS SPECIFICATIONS
The project uses visualizations to analyze which fields have a high amount of missing/null values or high cardinality.	<p>The project correctly identified which field(s) has/have a high amount of missing/zero values.</p> <p>The project correctly identified which field(s) has/have a Gaussian distribution shape based on the histogram analysis.</p> <p>The project correctly identified fields with high cardinality.</p> <p>The project justified why these fields had high cardinality.</p> <p>The project correctly describes the distributions for age and gender.</p> <p>Optional: The project uses Tensorflow Data Validation Visualizations to analyze which fields have a high amount of missing/null values or high cardinality.</p>

CRITERIA	MEETS SPECIFICATIONS
The project selects features based on exploratory data analysis.	<p>The project correctly identifies whether to include/exclude payer_code and weight fields.</p> <p>The project justified why these fields should be included/excluded by using supporting data analysis.</p>

Data Preparation

CRITERIA	MEETS SPECIFICATIONS
The project uses the correct level(s) for the given EHR dataset (line, encounter, patient) and transforms, aggregates and filters appropriately.	The project uses the correct level(s) for the given EHR dataset (line, encounter, patient) and transforms, aggregates and filters appropriately.
The project correctly reduces dimensionality for a dataset containing a	The project correctly maps NDC codes to generic drug names and prints out the correct mappings in the notebook.

high cardinality code set field.	MEETS SPECIFICATIONS
The project dataset has been split correctly for EHR machine learning models.	<p>The project has correctly split the original dataset into train, validation, and test datasets.</p> <p>The projects dataset splits do not contain patient or encounter data leakage.</p> <p>The Projects code passes the Encounter Test.</p>

Feature Engineering

CRITERIA	MEETS SPECIFICATIONS
The project correctly creates categorical features using Tensorflow's feature column API for transforming the data.	<p>The project correctly completes the categorical feature transformer boilerplate function.</p> <p>The project successfully uses this function to transform the demo dataset with at least one new categorical feature.</p>

CRITERIA	MEETS SPECIFICATIONS
<p>The project correctly creates numerical features using Tensorflow's feature column API for transforming the data.</p>	<p>The project correctly completes the numerical feature transformer boilerplate function.</p> <p>The project successfully uses this function to transform the demo dataset with at least one new numerical feature.</p> <p>The project's transformer function correctly incorporates the provided z-score normalizer function for normalization or another custom normalizer.</p>

Model Building and Analysis

CRITERIA	MEETS SPECIFICATIONS
<p>The project correctly utilizes the regression model predictions for uncertainty estimation and classification output analysis.</p>	<p>The project has prepared the regression model predictions for TF Probability and binary classification outputs by doing the following:</p> <ul style="list-style-type: none"> Correctly utilized TF Probability to provide mean and standard deviation prediction outputs Created an output prediction dataset that has the labels correctly mapped to a binary prediction and actual value.

CRITERIA	MEETS SPECIFICATIONS
The project correctly evaluates the model predictions across key classification metrics.	<p>The model has been evaluated across the following classification metrics: AUC, F1, precision, and recall.</p> <p>Students have completed both questions for the model summary and address bias-variance tradeoff in regard to this problem.</p>
The project utilizes the Aequitas toolkit to correctly create a bias report for race and gender on a provided dataset with some justification for their analysis.	<p>The project contains a bias report with the following:</p> <ul style="list-style-type: none"> • A visualization of at least two key metric(s) for patient selection • A visualization showing at least one reference group fairness example and its comparison on at least one metric (e.g. TPR). • Justification for analysis made about at least one visualization

Suggestions to Make Your Project Stand Out!

1. Students can tune the TF Probability library to show and rank uncertainty estimates.
2. Students can use cross-feature columns.
3. Students can use ICD groupings to reduce dimensionality.

4. Students can build a more complex representation of the data - e.g. conditional probability matrix, pre-trained embeddings, etc.
 5. Students uses Tensorflow Data Validation Visualizations to analyze which fields have a high amount of missing/null values or high cardinality.
-
-