TEMPUS

Data Science Case Study

Background

At Tempus, we are continually evaluating our data to see how it can be clinically informative. In order to better assess your skills and creativity, we would like you to provide examples of your data inspection, analyses, and modeling choices on the accompanying data set.

The data is synthetic yet intended to resemble the kind of data and modeling challenges frequently encountered by the Tempus data science team.

Assignment

The assignment is to investigate and analyze the accompanying data set, producing a model that can be used to predict the “target”.

The data set consists of three files:

- *targets.csv* - contains the binary “target” labels for the data set to be used in modeling, matched to a pair of unique identifiers for a patient’s profile and biomarkers data
- *patient_profiles.json* - contains various information on patient data points
- *biomarkers.csv* - contains several binary columns marking the presence of known biomarkers (though, deliberately left without further context)

The information present in *patient_profiles.json* and *biomarkers.csv* are to be used to train and predict the labels in *targets.csv*.

Deliverables

- A written explanation of what work you did on the assignment and why
- The code you used to perform the work (*Note*: Feel free to combine your written explanations alongside you code, such as in the form of a code notebook)
- A description of the predictive model produced along with a discussion of how well the model performs

Expectations

There is no such thing as a perfect model, only what can be done in a certain amount of time. We do not expect you to work on this task for more than 4 hours. We value your thought process, communication, and creativity just as well as how the model performs.