FHIR Features: Training Machine Learning Models to Understand FHIR

FHIRFLY · Follow
3 min read · Jul 18, 2023

Listen Share More

Healthcare has become more digital and data-driven than ever before. Data interoperability — the ability to share and make use of information across different health information systems — is a critical aspect of modern healthcare. One of the frameworks at the forefront of healthcare data interoperability is Fast Healthcare Interoperability Resources (FHIR).

FHIR, developed by HL7, is a standard for exchanging healthcare information electronically. It leverages the power of modern web technologies to provide a robust, extensible data model with a built-in mechanism for extensions and adaptations. In this article, we'll focus on a crucial part of FHIR — the Patient resource — and discuss how machine learning models can leverage it to make predictive analytics.

FHIR Patient Resource: A Closer Look

The Patient resource, as defined in FHIR, is used to represent information about an individual receiving healthcare services. It includes a wide range of demographic and clinical data, including but not limited to:

- Identifier: Unique identifiers assigned to the patient, such as a patient number or medical record number.
- Name: The name(s) associated with the patient, including given, family, and other names.

- Gender: The patient's gender (male, female, other, or unknown).
- BirthDate: The patient's date of birth, which is critical for age-related calculations.
- Address: The patient's home and mailing address.
- Contact: Contact details for the patient, including home, work, and emergency contact information.
- Communication: The patient's preferred language.

This wealth of demographic information can serve as inputs for machine learning models, providing a foundation for predictions related to patient health outcomes, utilization of healthcare services, and more.

Training Machine Learning Models with FHIR Patient Resource

One of the beautiful things about FHIR is its ability to standardize and structure data. This structured data can be directly fed into machine learning models, bypassing much of the tedious data cleaning and wrangling processes often required with healthcare data.

Let's dive into how we can use the features of the Patient resource to train machine learning models:

Data Preprocessing

Before using the data, we need to ensure it is in a format suitable for machine learning models. This usually involves data cleaning and encoding categorical variables into numerical ones. For instance, gender can be one-hot encoded to create binary "male", "female", "other", and "unknown" variables. Dates, such as birth dates, may be transformed into age at a given point in time or categorized into age groups. Addresses could be geocoded to provide spatial features such as urban vs. rural or socioeconomic indicators.

Feature Selection

Feature selection involves identifying the most important features for the machine learning model. This might be done through statistical methods, such as correlation or mutual information, or using machine learning techniques like feature importance from decision trees. For example, age and gender might be important features for

predicting certain health outcomes, while other features like preferred language might be more critical for predicting healthcare service utilization.

Model Training

After preprocessing the data and selecting relevant features, you can use this data to train your machine learning model. This process involves feeding the data into the model, allowing it to learn patterns and relationships among the features.

Model Evaluation

Once the model has been trained, it's essential to evaluate its performance using appropriate metrics. These could include accuracy, precision, recall, F1 score, or area under the receiver operating characteristic curve (AUC-ROC), depending on the specific application and type of prediction.

Moving Forward

The FHIR Patient resource is a treasure trove of information that can drive insightful predictive models in healthcare. However, it's important to remember that ethical considerations are paramount when working with healthcare data. Always ensure that patient data is anonymized and used responsibly, complying with regulations such as the Health Insurance Portability and Accountability Act (HIPAA) and the General Data Protection Regulation (GDPR).

FHIR and machine learning together have the potential to revolutionize healthcare, enabling personalized medicine and predictive healthcare at an unprecedented scale. By understanding and effectively utilizing the features of FHIR, we can unlock the full potential of machine learning in healthcare.

As we move into the future of digital health, it's clear that FHIR is not just an essential component of healthcare data interoperability, but also a vital resource for building intelligent, data-driven solutions in healthcare.

Fhir | Machine Learning | Healthcare | Data Science | Kaggle