Model Card for Predicting A&E Attendance from Primary Care Data

Inputs

**Demographic Data**  
Gender, Age, Is BAME (Black, Asian, Minority Ethnicity), Actual Ethnicity (e.g., Black British), Is Housebound (Boolean).

**Clinical Data**  
Presence of long-term conditions (e.g., Coronary Heart Disease, Hypertension), Number of Long-Term Conditions, Latest Observational Values (BMI, eGFR, Cholesterol, Hba1c).

**Socioeconomic Data**  
Deprivation Index, Health Index.

**Risk Factors**  
Risk Score from in-house algorithm,

Outputs

**No of A&E Attendances**  
A continuous variable representing the frequency of A&E attendances in the last 3 years.

**A&E Attendance in the last 3 years (Boolean)**  
 A binary classification indicating whether the patient attended A&E in the last 3 years (Yes/No).

Model Architecture

**Logistic Regression**  
To classify whether a patient is likely to attend A&E in the last 3 years (binary classification).

**Random Forest**  
 To predict the frequency of A&E attendances (regression) and to identify key determinants of A&E attendance.

Performance

There were no performance issued with creating the final model as the dataset used was relatively small (4500 records).

Limitations

**Limited Generalisability**  
The dataset is derived from a single GP practice in South West London, which may not reflect the demographics or healthcare patterns of other regions in the UK.

**No Data on Primary Care Access**  
Factors such as access to primary care and patient confidence in their GP are not included, which could be significant predictors of A&E attendance.

**Temporal Data Lag**  
The A&E data is 4-6 weeks out of date, which may result in minor inaccuracies for recently registered patients.

**No Hospital Admission Data**  
The dataset does not include information on whether A&E attendances led to hospital admissions, which could provide additional context for risk stratification.

Trade-offs

**Accuracy vs. Interpretability**  
Logistic regression provides interpretable results but may sacrifice some predictive accuracy compared to more complex models. Random Forest improves accuracy but is less interpretable.

**Model Complexity vs. Overfitting**  
Random Forest can capture complex relationships but risks overfitting, especially with a relatively small dataset (4,495 patients).

**Resource Allocation vs. Predictive Power**  
The model may identify high-risk groups for targeted interventions, but this could divert resources from other areas of need, potentially creating inequities.

This model card will be updated as the project progresses and more data becomes available.