Project Documentation

1. Data Loading  
   The Patient\_Demo.csv, and Patient\_Hospital\_Visit.csv are imported into python using the pandas.read\_csv method.  
     
   The two tables are merged on the institution\_id, and patient\_id to ensure data consistency.   
     
   The dates columns have been converted to datetime format using the pandas.to\_datetime method.  
   nullvaues where checked for using the isnull().sum() method and null values were found in admitted\_at, and discharged\_at respectively.

For the continuous columns with null values, the distribution was plot. For columns with gaussian distribution, the missing values were filled with the mean. For skewed distribution, the missing values were filled with the median since median is less sensitive to outliers compared to mean.

1. Feature Engineering

The age datapoint have been obtained from inserted\_at (which is the visit date) and dob (date of birth).  
The data has been aggregated by the number of visits per insititution, average age, how many states each institution is present.

All these aggregated features were merged into a single dataframe. Temporal features like day of the week, month, day of year, 1 day lag, 7 days lag, 7 days rolling mean, 30 days rolling mean, 7 days standard deviation, and facility type were also adopted.

The missing values that resulted from the lag values were filled with 0.

Correlation Plots were adopted to see the relevance of each feature to the target. Also, it is necessary to avoid multicollinearity. The final features adopted include day of the week, month, day of year, lag\_1, lag\_7, rolling\_mean\_7, rolling\_mean\_30, and rolling\_std\_7, and facility type.

1. Resource Demand Prediction

The data scaled using standard scaler. Timeseries method was adopted in forecast the daily resource demand for the next one month. The data was split into train and test. The data was trained from January 2021 to November 2021 and tested in December 2021. Various models were adopted but the two models that gave the best results were SARIMAX and Xgboost. SARIMAX was a strong fit because of its ability to capture seasons and trends, and Xgboost because of its ability to model complex relationships between the variables and the target.

The results from the text data were compared with the predicted results for both models. This was achieved via plots and recording the mean squared error, and the root mean squared error. They are about 88 distinct institution, therefore, to keep track of the predictions only the results from the first 10 institutions were displayed.

The performance of the SARIMAX model is displayed below:

|  |  |  |
| --- | --- | --- |
| Institution\_ID | MAE | RMSE |
| 09415354-4eb1-4866-87bb-efffe23e196a | 2.911506 | 4.403445 |
| 097d4418-6d7c-4922-a703-bde5a6b64919 | 6.747774 | 9.032736 |
| 0da8e9a5-855d-42c5-9435-d69e1019cafc | 1.870326 | 2.326755 |
| 0fa2e546-c292-4060-a05e-beb53fc31e72 | 2.917434 | 3.618631 |
| 18cb6e95-6073-4102-989e-1b133ac99b29 | 5.133061 | 6.615361 |
| 19adebda-8361-4267-8ab0-7fe734377602 | 2.059718 | 3.355206 |
| 19ae2f00-60e7-4db1-8719-644820cc04df | 4.575017 | 6.046701 |
| 1a75706d-889a-48ab-a6c7-1fbbd70aeda8 | 1.10079 | 1.418022 |
| 23533011-6b2e-4aaf-a992-85fcf7bf04d9 | 0.524366 | 0.641303 |
| 2fa0d571-e424-4ee7-ac8b-deb7792305a7 | 5.095736 | 6.647599 |

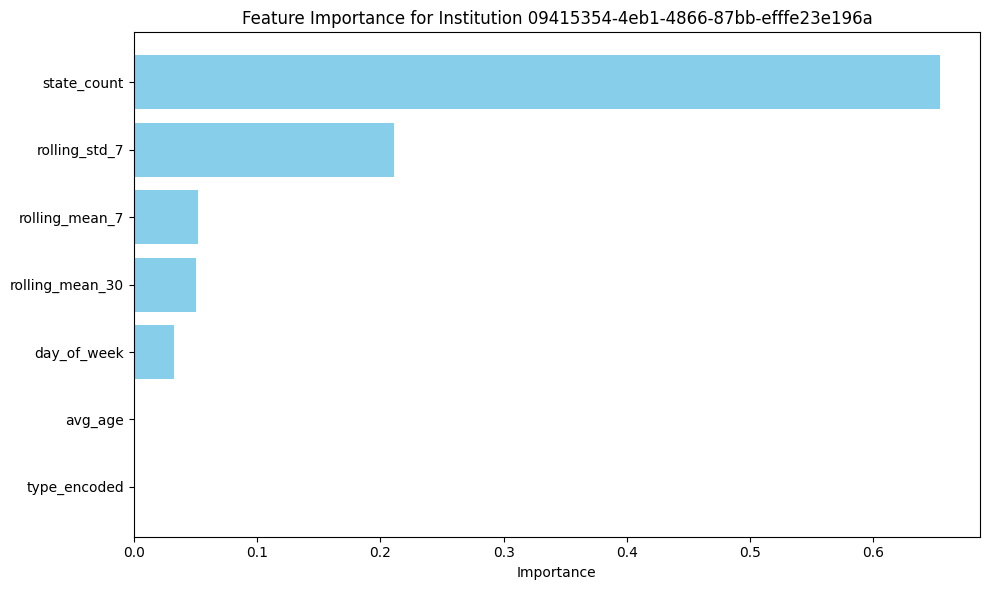
The performance of the Xgboost model is displayed below:

|  |  |  |
| --- | --- | --- |
| Institution\_ID | MAE | RMSE |
| 09415354-4eb1-4866-87bb-efffe23e196a | 4.465753 | 6.727296 |
| 097d4418-6d7c-4922-a703-bde5a6b64919 | 8.664539 | 10.660311 |
| 0da8e9a5-855d-42c5-9435-d69e1019cafc | 2.348906 | 2.812373 |
| 0fa2e546-c292-4060-a05e-beb53fc31e72 | 2.835082 | 3.768774 |
| 18cb6e95-6073-4102-989e-1b133ac99b29 | 6.507390 | 8.080336 |
| 19adebda-8361-4267-8ab0-7fe734377602 | 2.161287 | 3.443570 |
| 19ae2f00-60e7-4db1-8719-644820cc04df | 4.654722 | 6.452037 |
| 1a75706d-889a-48ab-a6c7-1fbbd70aeda8 | 1.336590 | 2.218859 |
| 23533011-6b2e-4aaf-a992-85fcf7bf04d9 | 0.079900 | 0.455083 |
| 2fa0d571-e424-4ee7-ac8b-deb7792305a7 | 6.002909 | 8.245859 |

Resource Allocation  
python optimization libraries like pulp has been used to predict the likely resource allocation per each institution and total visits.

1. Model Interpretability and Explainability

The features the model deemed to be the most important when making predictions have been obtained using feature importance of the Xgboost model. The plot is provided below.



From the plot above, the notable driver of volume of visitation is the number of states institutions have their presence. The more branches they have, the more likely the increase in the volume of visitations for these institutions.

1. Deployment and Monitoring  
   Using Epic EMR as a case study.

**Understand Epic’s Architecture and Tools**

* **Epic App Orchard**: This is Epic’s platform for third-party app integration. It provides APIs, documentation, and resources for developers to build and integrate tools.
* **FHIR APIs**: Epic supports Fast Healthcare Interoperability Resources standards, making it possible to access and integrate with EMR data securely.
* DataWarehouse: Using Epic’s Cogito (data warehouse) analytics suite to extract historical data for model training and validation.

**Model Deployment Steps**

**a. Design the Integration Workflow**

* **Real-Time Prediction**:
  + Using Epic'sFHIR APIs to extract relevant data (e.g., patient vitals, demographics, clinical notes) in real time.
  + Processing the data with the pickled SARIMAX or Xgboost model.
  + Send predictions back to Epic using FHIR resources like Observation or ServiceRequest.

**b. Set Up Infrastructure**

* Host the model externally on:
  + Cloud Platforms: Azure endpoint.
  + On-Premise: If required by compliance policies, deploy the model within the healthcare provider's data center.
* Using Epic Bridges for data exchange between the EMR and the external model service.

**c. Embed Predictions in Epic**

* **Dashboards**: widgets can be created within Epic Hyperspace to display predictions to clinicians.
* **Alerts**: notifications can be triggered based on model predictions through Epic's In-Basket or workflow engine.
* **Decision Support**: The results can be integreated into Epic's Clinical Decision Support (CDS) to provide actionable insights during the care process.

**Monitoring and Maintenance**

1. **Monitor Model Performance**

* Tools like Grafana can be used for monitoring metrics like latency, prediction accuracy, and system usage.

**b. Anomaly Detection**

* Logging can be used to detect anomalies from input data and predictions.
* For real-time alerts, Epic’s RulesEngine can be used to capture abnormal trends in predictions or input data.

**c. Feedback Loop**

* Feedback from clinicians can be captured through Epic’s UI to understand how predictions are used and identify areas for improvement.

**Model Updates**

**a. Data Pipelines**

* Utilize Epic’s Cogito Clarity (for SQL-based data extraction) to source updated training data.

**b. Retraining Models**

* Automate pipelines to update the model periodically as new EMR data becomes available. Leverage tools like MLflow for managing retraining workflows.

**Security and Compliance**

* Ensure strict HIPAA compliance for all patient data interactions.
* Use OAuth 2.0 for secure authentication with Epic’s APIs.
* Log all model usage and interactions for auditability, using Epic’s logging tools or external systems.

**Explainability and Usability**

* **Explainability**: Use SHAP to create user-friendly explanations of predictions that can be embedded in Epic’s interface.
* **Clinician Training**: Host training sessions to familiarize staff with how predictions are generated and interpreted.
* **Customization**: Collaborate with stakeholders to tailor model outputs to fit specific clinical workflows.

**LIMITATIONS**

The data on the project on the resource allocation optimization lacks general resources like beds e.t.c. These resources are needed by the optimization model to assign resources to the existing institutions and the forecasted resource demand.