Three daily staffing hour variables, namely Total Hours for LPN (measured in Hrs_LPN), Total Hours for RN (measured in Hrs_RN), and Total Hours for CNA (measured in Hrs_CNA), will be considered as the response variables for the prediction models. The performance of the models will be evaluated using three metrics, namely Mean Absolute Error (MAE), Mean Absolute Percentage Error (MAPE), and Mean Interval Score (MIS). For the kth caregiver type (e.g., CNA, LPN and RN), the performance metrics can be explicitly written as:

$$\text{MAE}_k = \frac{1}{|S| \cdot |T|} \sum_{\forall i \in S} \sum_{\forall t \in T} \text{abs}(y_{k,i,t} - \hat{y}_{k,i,t})$$

$$\text{MAPE}_k = 100 \times \frac{1}{|S| \cdot |T|} \sum_{\forall i \in S} \sum_{\forall t \in T} \text{abs}(\frac{y_{k,i,t} - \hat{y}_{k,i,t}}{y_{k,i,t}})$$

$$\text{MIS}_k = \frac{1}{|S| \cdot |T|} \sum_{\forall i \in S} \sum_{\forall t \in T} \left((u_{k,i,t} - l_{k,i,t}) + \frac{2}{\alpha}(l_{k,i,t} - y_{k,i,t})I(y_{k,i,t} < l_{k,i,t}) \right.$$
$$\left. + \frac{2}{\alpha}(y_{k,i,t} - u_{k,i,t})I(y_{k,i,t} > u_{k,i,t}) \right)$$

where S and T are index sets representing NHs and time points to be predicted, respectively, and $|\cdot|$ is the cardinality of a set. $y_{k,i,t}$ is the observed total staffing hours for caregiver type k in NH i on day t, $\hat{y}_{k,i,t}$, $u_{k,i,t}$ and $l_{k,i,t}$ are the predicted total staffing hours of caregiver type k in NH i on day t, its upper and lower prediction bounds, respectively, with a significance level α=0.05. $I(\cdot)$ and abs($\cdot$) are indicator function and absolute value function, respectively .

Training Data: Daily staffing hour data for LPN, RN, and CNA (i.e., Hrs_LPN, Hrs_RN, and Hrs_CNA) from 04/01/2024 to 06/30/2024 for 100 existing NHs along with basic organizational characteristics.

Phase 1 Prediction: Based on the training data provided, the team needs to develop prediction models to forecast Hrs_LPN, Hrs_RN, and Hrs_CNA between 04/01/2024 to 06/30/2024 for NHs in each of the following 20 categories: 2 different state locations, 5 different capacity levels, and 2 different quality rating levels. The judges will randomly select 20 new NHs from these 20 categories to evaluate the prediction performance of the developed model. In phase 1, the NHs index set S_1 consists of 20 new NHs. The time points index set T_1 consists of all days from 04/01/2024 to 06/30/2024.