adversarial attacks on battery energy storage systems (BESS):

1. Model Selection and Justification:
   * In my research, I used models specifically tailored to adversarial attack detection and simulation on BESS. Rather than traditional time-series forecasting models like ARIMA, I employed physics-based models (such as PyBaMM) for better representation of the internal battery dynamics. These models can simulate the impact of adversarial attacks on current, temperature, and voltage in a more detailed way, capturing real-world interactions that purely statistical methods might miss. The added advantage of PyBaMM is that it allows for adversarial scenarios, which traditional methods like ARIMA cannot handle.
2. Feature Engineering:
   * The key features in my research were current, temperature, voltage, and attack windows. These were engineered based on my attack scenarios where each sensor could be exploited. Historical SoC and voltage values, combined with noise injections, provided insights into the degradation process during the attack. Feature selection was not as typical as in standard ML models; rather, I chose features that directly affect battery degradation and performance during attacks. Through detailed simulations, I observed how these features contributed to voltage collapse and overall system destabilisation.
3. Handling Data Issues:
   * For data quality, I carefully managed sensor noise by simulating small noise perturbations in the model to reflect real-world operational noise in BESS. Missing data was less of an issue due to the controlled nature of the simulations, but if encountered, linear interpolation would be used, especially for critical variables like voltage or current.
4. Temporal Dependencies:
   * My model captures both short-term and long-term dependencies using attack windows designed to mimic sudden sensor malfunctions and gradual system degradation. By setting specific start and end times for adversarial attacks, I ensured the model could reveal short-term spikes (e.g., current surges) and long-term consequences (e.g., voltage collapse), allowing an understanding of how temporal disruptions in sensor data affect overall system stability.
5. Evaluation Metrics:
   * I used metrics like voltage deviation, current spikes, and SoC/SoH deterioration over time to assess model performance. Traditional metrics like RMSE or MAE were less applicable in my context, where the focus was more on observing the system's deviation from normal conditions during an attack. For example, the sharp voltage drop after 400 seconds provided a clear indicator of attack success.
6. Hyperparameter Tuning:
   * In my case, hyperparameter tuning was mostly about adjusting attack windows and noise magnitudes. I experimented with different levels of perturbations in the temperature and current to see the greatest impact on system destabilisation. The challenge was balancing attack intensity with the need to maintain a stealthy approach—too much noise would make the attack easily detectable, while too little wouldn't disrupt the system sufficiently.
7. Model Interpretability:
   * Given the complexity of adversarial scenarios, interpretability was enhanced by focusing on key system-level metrics like voltage and current behaviour over time. The voltage collapse curve provided a direct measure of system failure, while the sharp temperature spikes indicated overheating due to the attack. Rather than using methods like SHAP or LIME, the physical behaviour of the system during attacks was self-explanatory in revealing the consequences of adversarial interference.
8. Seasonality and External Factors:
   * The focus of my research was less on external environmental factors like seasonality, as it was more geared toward exploring internal vulnerabilities of BESS. However, the impact of temperature on system performance was highlighted, where spikes in temperature significantly destabilised the system. This demonstrates that even without external seasonality, internal temperature fluctuations can have devastating effects on battery performance.
9. Overfitting Prevention:
   * Overfitting was controlled by using real-world operational noise and attack scenarios to generate data, ensuring that my models were not just tuned to one specific attack pattern. This way, the model remains robust to different levels and types of adversarial attacks. The use of general physical models like PyBaMM also prevents overfitting to particular datasets, as these models represent fundamental electrochemical processes rather than specific data trends.
10. Comparative Analysis:
    * My approach demonstrates significant advantages over simpler models or statistical methods. While basic forecasting techniques may predict voltage or current trends, they fail to capture the compounded effects of adversarial attacks on multiple sensors. My combined temperature and current attacks produced a clear deviation in voltage collapse, which standard models wouldn't detect.
11. Scalability and Deployment:
    * My model is scalable and could be integrated into real-time energy management systems, especially for anomaly detection in grid-scale batteries. The computational requirements are manageable with modern hardware, and the use of PyBaMM makes it possible to simulate these attacks in various battery configurations, allowing for broader applications across different battery technologies.
12. Model Adaptability:
    * My model adapts well to different types of adversarial attacks and noise levels. By varying attack windows and noise intensities, I demonstrated that the system could simulate different levels of sensor malfunction and external disruption. Though I didn’t specifically apply transfer learning, the adaptability of the model comes from its ability to simulate various attack scenarios, rather than relying on pre-learned data distributions.