In Class Activity
Classification

HARSH SIDDHAPURA
1230169813

03/21/2024

1. **Underrepresentation of Sample:** When certain groups or categories are inadequately represented in a dataset, it can lead to biased or incomplete insights. For instance, consider a healthcare study aiming to predict the effectiveness of a new treatment. If the study primarily includes patients from a specific age group or ethnicity, the resulting model may not generalize well to the broader population.
    - ***Impact on Accuracy:*** The model's accuracy suffers because it lacks exposure to diverse scenarios. It may make erroneous predictions for underrepresented groups.
    - ***Example***: A skin cancer detection model trained predominantly on fair-skinned individuals may struggle to accurately identify melanomas in people with darker skin tones.

2. **Not Capturing All Relevant Attributes:** In complex problems, omitting critical features can hinder accurate predictions. Imagine building a fraud detection system for credit card transactions. If the model only considers transaction amounts and timestamps but ignores contextual information (e.g., location, user behavior), it will miss important cues.
    - ***Impact on Accuracy:*** The model's predictions become myopic, overlooking crucial context. Accuracy suffers due to incomplete information.
    - ***Example***: A recommendation engine for streaming services that ignores user preferences (such as genre preferences or viewing history) will struggle to suggest relevant content.

3. **Noise in the Data:** Noise refers to random fluctuations or errors present in the dataset that do not carry meaningful information. These can arise due to measurement inaccuracies, sensor malfunctions, or other external factors.
    - ***Impact on Accuracy:*** When noise contaminates the data, it misleads the model during training. The model may mistakenly learn from these irrelevant patterns, leading to suboptimal predictions.
    - ***Example***: Imagine a climate prediction model that incorporates temperature data from faulty sensors. If occasional spikes or erroneous readings occur, the model might associate them with specific weather conditions, affecting its accuracy.

4. **Patterns Beyond Model Capabilities:** Some real-world phenomena exhibit intricate, nonlinear relationships that surpass the modeling capabilities of simple algorithms. Linear models, for instance, struggle to capture complex dynamics.
    - ***Impact on Accuracy:*** Oversimplified models fail to represent reality accurately. Their predictions deviate significantly from actual outcomes.
    - ***Example***: Consider predicting housing prices based solely on square footage using linear regression. However, housing values depend on various factors (location, amenities, market trends) that linear models cannot adequately capture. Consequently, the model's predictions will be inaccurate.