**5.3 Assignment**

**Model Validation and Verification Reflection**

For this assignment, I walked on the video tutorials where it is needed for it to be focused on these core areas of these machine learning models which is required to be

1. **“Cross-Validation in Machine Learning”** by StatQuest with Josh Starmer
2. **“Confusion Matrix, Accuracy, Precision, Recall, F1 Score”** by Krish Naik
3. **“Bias and Fairness in AI”** by MIT CSAIL

**Key Takeaways:**

From **StatQuest’s video**, I have learnt for the cross-validation for the **k-fold cross-validation**, improves model reliability. By splitting the dataset into *k* parts and rotating which fold is used for validation, this technique helps detect overfitting and provides a more stable estimate of performance. It ensures that every observation is used for both training and validation.

Krish Naik’s video broke down the **confusion matrix** and its derived metrics. Accuracy is often not enough. This is needed to be a part of **precision, recall, and F1-score** provide a clearer picture, especially with imbalanced datasets. For example, in a medical diagnosis model, **high recall** ensures most true positives (sick patients) are caught, even at the cost of some false alarms.

The MIT CSAIL lecture highlighted real-world ethical concerns such as **algorithmic bias**, especially when training data reflects societal inequalities. The video stressed the need for **transparency**, **bias audits**, and involving diverse stakeholders during model development.

**Validation Techniques:**

* **Train-Test Split**: Divides data into two sets (e.g., 80/20) to evaluate how well the model generalizes. It’s quick but can give unstable results if the dataset is small.
* **Cross-Validation**: Particularly k-fold CV, is more robust as it tests the model on multiple data partitions, reducing variance in performance estimates.

**Performance Metrics:**

* **Accuracy** is the overall correctness but can be misleading with class imbalance.
* **F1-score** balances **precision and recall**, making it ideal for applications like medical diagnosis or fraud detection where both false positives and false negatives carry serious consequences.

**Ethical Concerns:**

* **Bias in training data** can perpetuate discrimination. For instance, in loan default prediction, historical biases may unfairly disadvantage certain demographic groups.
* **Privacy** is critical, especially in healthcare. Ensuring patient data is anonymized and securely stored is both an ethical and legal necessity.

This assignment helped me grasp not only the **technical** but also the **human aspects** of machine learning. I now better understand the importance of choosing the right validation methods and being mindful of how these models impact real lives.