

ML: Assignment-1

Q.1.

- ① K-fold CV assumes independent and identically distributed data. It becomes misleading under dataset shift, where train and test distributions differ even though CV looks good.
- ② ② Time-aware validation, where the model is trained on past semesters and tested on future ones.
- ③ ③ Out-of-distribution holdout, where an entire next semester is kept unseen for testing.

Q.2.

- Random Forest usually outperforms bagging because it reduces correlation between trees by randomly selecting features at each split, whereas bagging often builds highly similar trees using the same dominant features. This lower correlation allows Random Forest to achieve better variance reduction and generalization.