Naïve Bayes Report

Preprocessing:

- The rating extraction in Naïve Bayes converts the Likert-scale responses into numerical values (1-5) using regex pattern matching
- Multi-select encoding processes task preference questions using MultiLabelBinarizer, which creates binary features for each task category (Math computations, Writing/debugging code, Data processing/analysis, Explaining concepts), so vectors like [0, 1, 1, 0, 1, 1, 1, 0]. This encoding choice is justified because it preserves independence between task categories, as needed by the assumption in Naïve Bayes
- Missing value handling drops rows with NaN values to maintain data quality

Prevent data leakage:

- Data is split 70/15/15
- During hyperparameter tuning, only training and validation sets are used to select the optimal alpha value (0.01)
- The test set remains completely untouched until final evaluation, accessed only once after all model selection decisions are finalized
- The MultiLabelBinarizer is fitted only on training data to prevent information about test set categories from influencing the encoding scheme

Validation method:

- Hold-out validation strategy with a 70/15/15 split (train/validation/test)

Optimization:

- For Naive Bayes, we tune the Laplace smoothing parameter alpha ($\alpha \in \{0.01, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0\}$) to balance between overfitting and underfitting
- No learning rate schedule is needed as Naive Bayes uses closed-form MLE

Hyperparameter list:

- Alpha (Laplace smoothing) $\in \{0.01, 0.1, 0.5, 1.0, 2.0, 5.0, 10.0\}$

Naïve Bayes Report

Hyperparameter choices:

- For Naive Bayes, alpha values span from 0.01 (minimal smoothing) to 10.0 (heavy smoothing)
- validation shows stable performance across this range (70% accuracy)
- The optimal α=0.01 was selected for it yielded the highest accuracy
    - The training data provides sufficient coverage without requiring aggressive smoothing

Evaluation metrics:

- In the training process, the Naïve Bayes model was able to utilize two other evaluation metrics: macro recall and macro F1