**Model Evaluation & Selection Report**

**Objective**

The primary goal was to perform model evaluation on a text classification task, comparing both traditional machine learning classifiers and deep learning models to determine the best-performing approach for the dataset.

**Models Evaluated**

| **Model** | **Accuracy** | **Precision** | **Recall** | **F1-Score** | **Loss** |
| --- | --- | --- | --- | --- | --- |
| SVC (Support Vector Classifier) | 1.0000 | 1.00 | 1.00 | 1.00 | 0.0000 |
| KNeighborsClassifier | 0.9881 | 0.99 | 0.99 | 0.99 | N/A |
| RandomForestClassifier | 1.0000 | 1.00 | 1.00 | 1.00 | 0.0000 |
| BERT (Fine-Tuned) | 0.3627 | 0.305 | 0.402 | 0.351 | 2.5001 |

**Model Selection Reasoning**

**Baseline Model:**

**BERT (Fine-Tuned)**

* BERT was selected as the baseline because it represents a modern transformer-based approach designed for text understanding.
* Despite its strength on large and complex datasets, it underperformed here (36% accuracy) due to the dataset's limited size, likely causing overfitting and suboptimal learning.

**Advanced Model Choice:**

**SVC (Support Vector Classifier)**

* Outperformed all other models with perfect classification results (100% accuracy, precision, recall, F1-score).
* SVC’s performance suggests that the problem space is likely well-separated in the feature domain and not complex enough to require deep contextual representations.
* Simpler models like SVC are better suited for small-to-medium sized datasets and lower computational budgets.

**Tuning Strategy**

* **BERT** was fine-tuned using standard pre-trained weights with early stopping; however, it couldn’t fully converge due to the dataset’s size.
* **SVC, RandomForest, and KNN** required minimal tuning:
  + Kernel type and regularization (SVC) were adjusted.
  + Number of neighbors (KNN) optimized for 5.
  + Number of estimators (RandomForest) set to default and proved sufficient due to dataset characteristics.

**Evaluation Outcomes**

* Classical ML models (SVC, RandomForest, KNN) achieved near-perfect performance.
* Deep Learning (BERT) struggled with this dataset, reinforcing that not every task benefits from over-complex architectures.
* Evaluation metrics, including **Accuracy, Precision, Recall, F1-Score, and Loss**, clearly favored SVC.

**Key Improvements Identified**

1. **Model Simplification**:  
   Simpler classifiers like SVC outperformed BERT for this task, reducing both computational cost and risk of overfitting.
2. **Dataset-Model Alignment**:  
   Classical ML models are more suitable when:
   * Data is tabular or feature-engineered.
   * The dataset is small to medium-sized.
   * The problem is clearly separable.
3. **Training Stability**:  
   Classical models showed more stable convergence and lower loss, indicating strong generalization even without heavy computational tuning.

**Conclusion**

For this task, classical machine learning models, particularly **SVC**, delivered the best results. Despite BERT’s capabilities in understanding context, the problem’s scale and complexity made SVC the more effective and efficient choice.  
Future improvements could include:

* Expanding the dataset.
* Using hybrid models combining engineered features with contextual embeddings.
* Applying cross-validation for further generalization checks.