Card Suits Classification Report

1. Data Analysis and Preprocessing

First, we checked the data for outliers by comparing the mean and median of each feature, and looked at the feature distributions with histograms. We used a Z-score threshold of 3.5 and the mean-median difference to spot outliers, removing 19 from the dataset. This reduced the dataset size from 1824 to 1805 samples.

Feature Selection

- Identified outliers using a Z-score threshold of 3.5 and mean-median difference in std units.
- Removed 19 outliers, reducing dataset size from 1824 to 1805 samples.
- Visualized feature combinations to select the best ones.
- Chose Features 2 and 4 for:

- **Clear class separation**
- Statistical stability
- Complementary information

- **Feature 2**: std = 0.0089
  Feature 4: std = 0.0010
- No normalization needed due to scale ratio of 8.9.

2. Classification Result

Base Performance

Classifier Error Rate

Independent: 2.58%

Multivariate: 0.71%

Parzen (h=0.0005): 1.75%

1-NN: 1.81%

Training Set Size Impact (5 repetitions)

Reduction	Independent	Multivariate	Parzen (h=0.0005)
10%	3.19% ± 0.42%	1.07% ± 0.26%	8.39% ± 1.46%
25%	3.85% ± 1.50%	0.77% ± 0.15%	5.76% ± 0.52%
50%	2.93% ± 0.50%	0.81% ± 0.02%	3.87% ± 0.23%

Prior Probability Approaches (Reduced Dataset)

Approach	Independent	Multivariate	Parzen (h=0.0005)
Without Prior	2.12%	0.37%	14.47%
Original Prior*	2.19%	0.37%	13.30%
Targeted**	2.79%	0.80%	13.56%

*Original: Black suits (0.165), Red suits (0.085)

**Targeted: Custom class-specific priors and part

Training Set Size Impact (5 Repetitions)

Reduction	Independent	Multivariate	Parzen (h=0.0005)
10%	3.19% ± 0.42%	1.07% ± 0.26%	8.39% ± 1.46
25%	3.85% ± 1.50%	0.77% ± 0.15%	5.76% ± 0.52%
50%	2.93% ± 0.50%	0.81% ± 0.02%	3.87% ± 0.23%

Per-Class Error Analysis

Reduction	Independent	Multivariate	Parzen (h=0.0005)
Without Prior	2.12%	0.37%	14.47%
Original Prior*	2.19%	0.37%	13.30%
Targeted**	2.79%	0.80%	13.56%

*Original: Black suits (0.165), Red suits (0.085)

**Targeted: Custom class-specific priors and parts

Per-Class Error Analysis

Class No	Prior	Original	Targeted
1	0.88%	0.88%	0.44%
2	0.00%	0.00%	1.75%
3	0.00%	0.00%	0.58%
4	0.00%	0.00%	0.00%
5	0.00%	0.00%	0.00%
6.	2.63%	2.63%	2.19%
7	0.00%	0.00%	0.88%
8	0.00%	0.00%	0.00%

Analysis Result

Classifier Performance performance

Base performance (full dataset): 0.71%

- **With data reduction:**
- 50% of data: 0.81% ± 0.02%
- 25% of data: 0.77% ± 0.15%
- 10%: 1.07% ± 0.26%

Maintains strong performance even with reduced data
Best performance across all congurations.
2. Independent Classier:
Base performance: 2.58% (full dataset)
Data reduction impact:

- 50%: 2.93% ± 0.50%
- 25%: 3.85% ± 1.50
- 10%: 3.19% ± 0.42%
  Consistent but higher error rates
  Most stable with full data set

3.**Parzen Classifier (h=0.0005):**

- **Base performance (full dataset):** 1.75%
- **Data reduction impact:**
  - 50% → 3.87% ± 0.23%
  - 25% → 5.76% ± 0.52%
  - 10% → 8.39% ± 1.46%
- **Summary:** Strong with the full dataset, but performance drops significantly with less data. Highly sensitive to dataset size.

4.1-NN Classifier:

- **Base performance:** 1.81% (similar to Parzen with the full dataset).
- **Summary:** Performs consistently well, on par with Parzen's full dataset performance.

Conclusion

1.Best Performing Configurations

Full Dataset Performance:

- **Multivariate:** 0.71% (best performer)
- **Parzen:** 1.75%
- **1-NN:** 1.81%
- **Independent:** 2.58%

Reduced Dataset Performance:

- **Multivariate**: Maintains strong performance
- **Parzen**: Significant performance drop
- **Independent**: Slight performance drop

2.Data size impact

Multivariate shows remarkable stability

Parzen requires full dataset for performance

Independent shows moderate degradation

3.Prior Probability Effects

Most effective with reduced dataset

Limited impact on full dataset performance

Can improve specific class performance

4.P**ractical Recommendations:**

- Use Multivariate classier for best overall performance
- Maintain full dataset when possible, especially for Parzen
- Consider computational trade-os vs. accuracy
- No need for feature normalization
- Use Parzen only with full dataset

The results indicate that while all classifiers yield good performance with the full dataset, the **Multivariate classifier** delivers the most consistent results across all configurations. The **Parzen classifier**, though competitive at 1.75% with the full dataset, requires careful attention to data size for optimal performance.