Executive Summary: Comparative Analysis of Support Vector Machine Performance Across Geophysics, Finance, and Medical Diagnostics

1. Purpose of Study

Checking patient biomarkers to verify the presence of the Hepatitis C Virus (HCV), predicting credit card defaulting within the banking sector, and developing an accurate system for earthquake alert prediction represent some of the real-world problems that can be resolved with the help of machine learning. This study proposes a comparative performance evaluation of the Support Vector Machine (SVM) across these three distinct domains. The aim is to test how well the SVM can handle different types of data and the specific challenges that each domain presents. By learning the intricate connections between features and labels, the model gains the necessary competence to successfully generalize its knowledge when facing new, real-world scenarios.

2. Methods to Gather Data

The research utilized three datasets: the Earthquake Alert Prediction Dataset (1300 samples, 6 seismic features), the Default of Credit Card Clients Dataset (30,000 credit card customers), and the HCV dataset (615 patients with biochemical data). Preprocessing involved excluding duplicate records, addressing missing values using mode imputation, managing outliers via class-wise median imputation, and applying standardization to all numerical features. The data was partitioned using an 80-20 ratio, Stratified K-fold Cross-Validation was employed for HCV Data, and the Support Vector Classifier was configured with the Radial Basis Function kernel with regularization parameter C=10 for Credit Card Default and C=25 for Earthquake Alert.

3. Findings

The SVM achieved exceptional performance in medical diagnostics (Accuracy: 0.9918) and solid results in geophysical classification (Accuracy: 0.8846), but showed significant limitations in financial risk assessment (Recall: 0.66). For the HCV Data, the model correctly identified 108 healthy patients and 14 diagnosed patients with only 1 False Negative, relying heavily on specific biochemical markers. The Earthquake Alert model successfully distinguished between four alert levels, with 64 red alerts correctly classified, using instrumental intensity (mmi) and significance (sig) as the most critical features. However, the Credit Card Default model struggled with the minority class, producing 841 False Negatives while correctly identifying 4,389 non-defaulters, resulting in low Precision (0.3515).

4. Report Limitations

Several factors limited this study. The HCV dataset's relatively small sample size and inherent class imbalance required specialized validation techniques. For the credit card default dataset, the severe class imbalance and the overlapping nature of financial behavior patterns created a challenging classification environment that the SVM could not totally address. This failure arises from the model's difficulty in handling noisy data where target classes exhibit overlapping properties in the feature space. The SVM is highly sensitive to the size and balance of the data, a core weakness of the soft margin optimization.

5. Conclusion

The experimental study showed a clear trade-off in the SVM's capabilities: the model showed high capability in medical diagnostics (Accuracy: 0.9918) and geophysical classification (Accuracy: 0.8846). However, the SVM demonstrated an essential weakness in financial risk assessment, achieving a low Recall (0.66) due to its strong tendency to struggle with both severe class imbalance and features that overlap significantly, making those target classes difficult to separate.