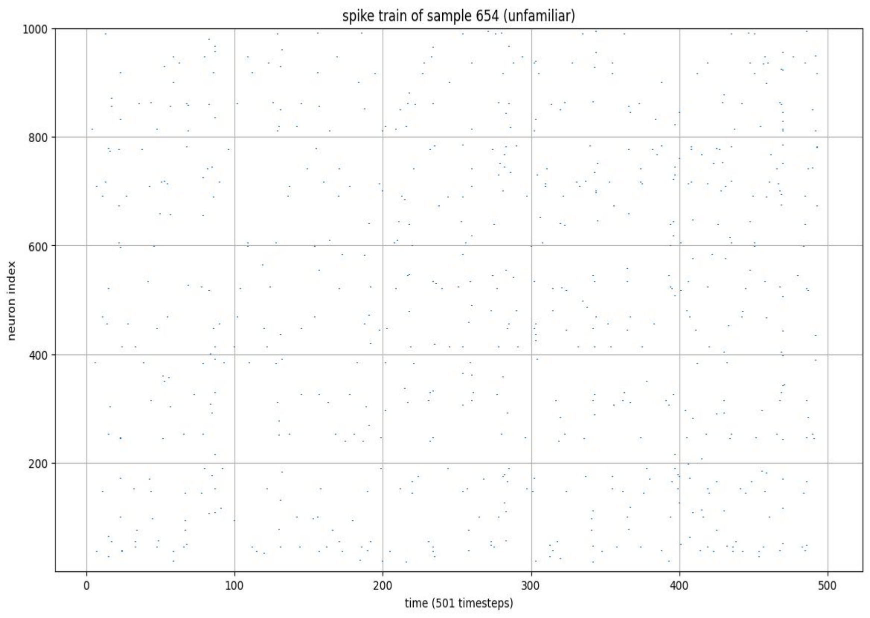
Neuroinformatics Project: Encoding Neural Data

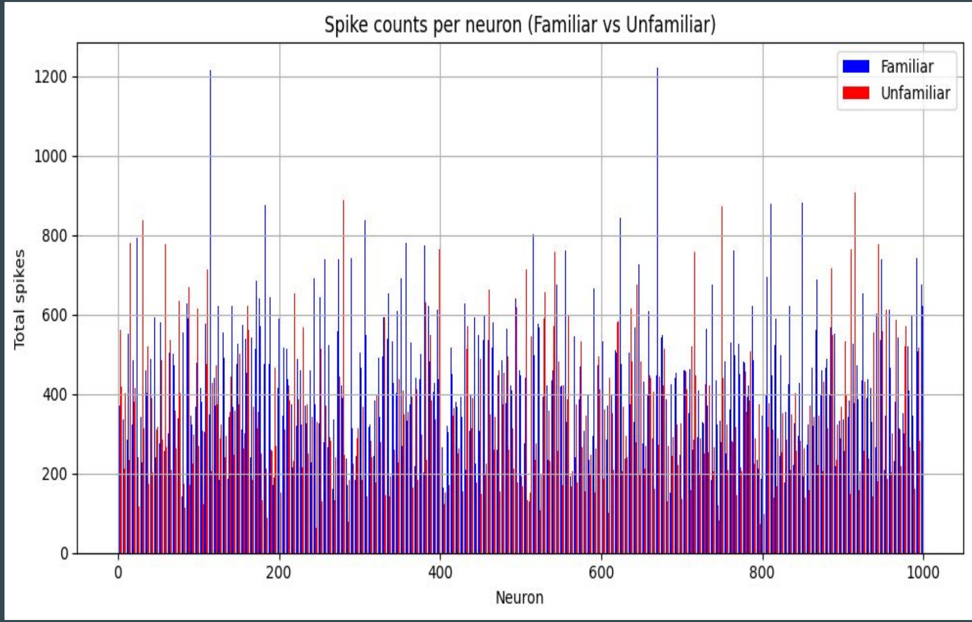
Silas Sinning WiSe24/25
Matrikelnummer: 996153

Overview

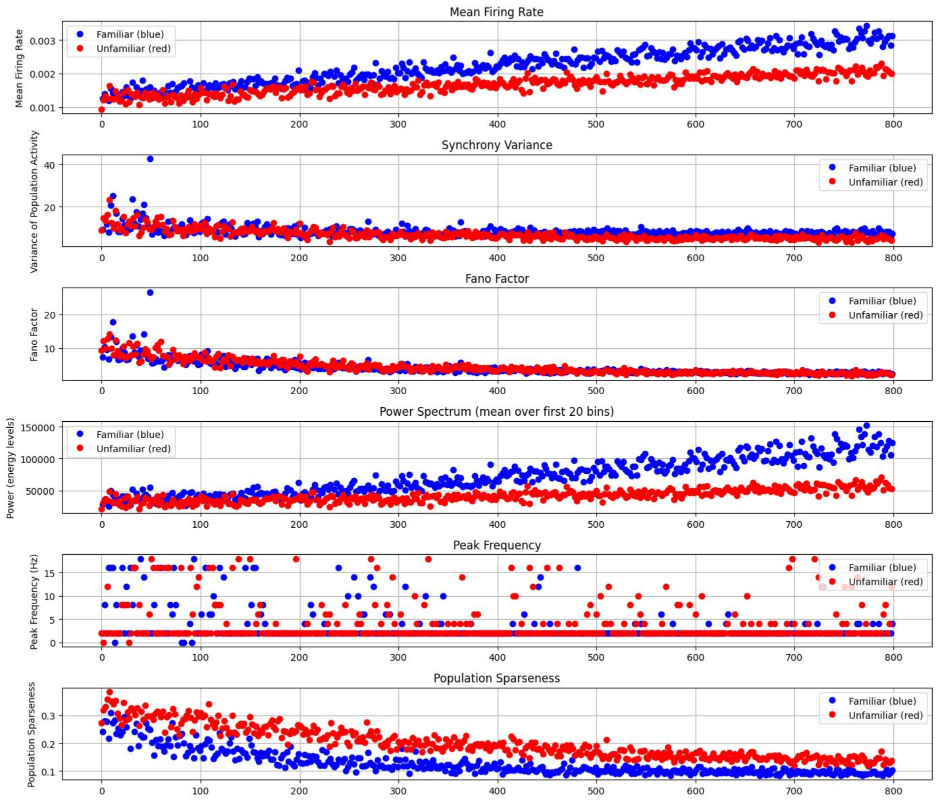
- Data Visualization
- Data Transformation and Exploratory Analysis
- Model Training and Classification
- Model Evaluation
- Model Comparison

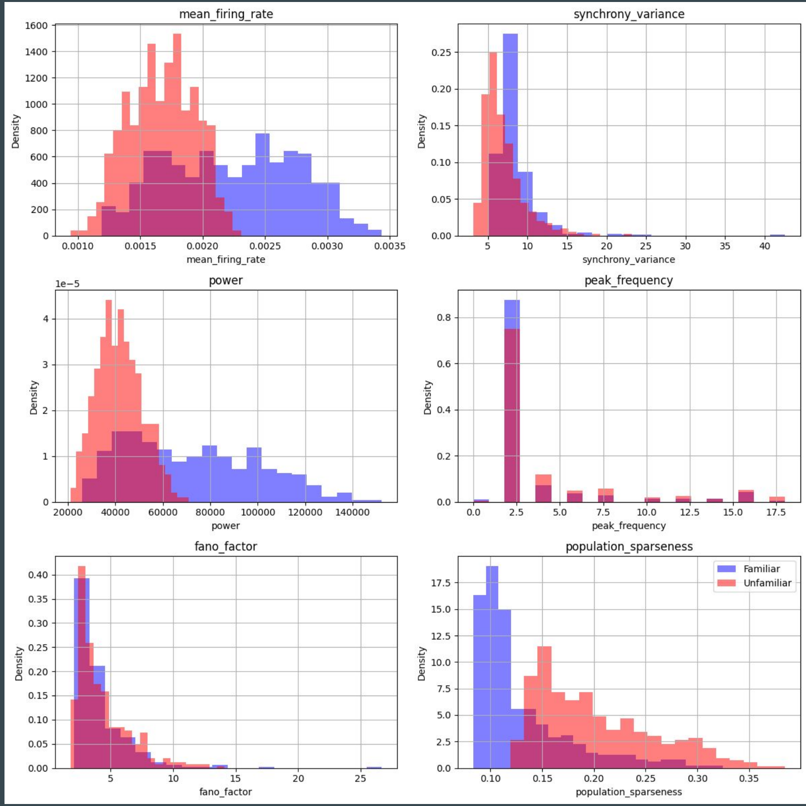
Data Visualization





Data Transformation and Exploratory Analysis





KL Divergence

Feature	KL Divergence
mean_firing_rate	13.66
synchrony_variance	0.52
power	5.30
peak_frequency	0.04
fano_factor	0.25
population_sparseness	14.73

Logistic Regression

$$P(y=1\mid x)=\sigma\begin{pmatrix} w_0 \\ +\,w_1\cdot \text{mean\_firing\_rate} \\ +\,w_2\cdot \text{synchrony\_variance} \\ +\,w_3\cdot \text{power} \\ +\,w_4\cdot \text{population\_sparseness} \end{pmatrix}$$

$$\left(\sigma(x)=\frac{1}{1+e^{-x}}\right)$$

- Models the probability of a binary outcome
- Applies a sigmoid to a linear combination of features
- Each weight shows how one feature shifts the log-odds

Model Evaluation

The model achieves an acceptable accuracy.

Data is largely linearly separable

The model is efficient, robust, relatively lean, interpretable and easy to implement.

- Average CV-Accuracy: 0.955 ± 0.027

- Confusion Matrix:
  [[88 1]
  [5 66]]

- F1- Score: 0.956522

- Precision Familiar Samples: 0.985075
- Precision Unfamiliar Samples: 0.946237

Model Comparison

k-Nearest Neighbors (k-NN)
- classifies samples based on the majority class among their k closest neighbors
- good for: data with clear local clusters and nonlinear boundaries

Random Forest
- based on decision trees
- good for capturing nonlinear patterns
- more human-interpretable but slower to train than linear models

Linear Discriminant Analysis (LDA)
- projects data onto a lower-dimensional space that best separates the classes
- good for well-separated classes and when features follow Gaussian distributions

Regularized Least Squares (Ridge Regression)
- linear
- good for high-dimensional data and collinear features

Model Comparison

	Model	Accuracy	F1 Score	Precision F	Precision UF
0	Random Forest	0.971875	0.969900	0.986395	0.959538
1	Poly Logistic Regression	0.968750	0.966887	0.973333	0.964706
2	LDA	0.965625	0.962963	0.986207	0.948571
3	Ridge	0.965625	0.962963	0.986207	0.948571
4	Logistic Regression	0.962500	0.959732	0.979452	0.948276
5	k-NN	0.959375	0.955631	0.992908	0.932961

Results

Random Forest outperforms all other models.

Polynomial Logistic Regression is in the second place showing that non-linear decision boundaries help, but not quite as much as Random Forest’s full flexibility

The still strong performance of Logistic Regression suggests that the data is largely linearly separable but there are subtle non-linear patterns that linear models can't fully capture