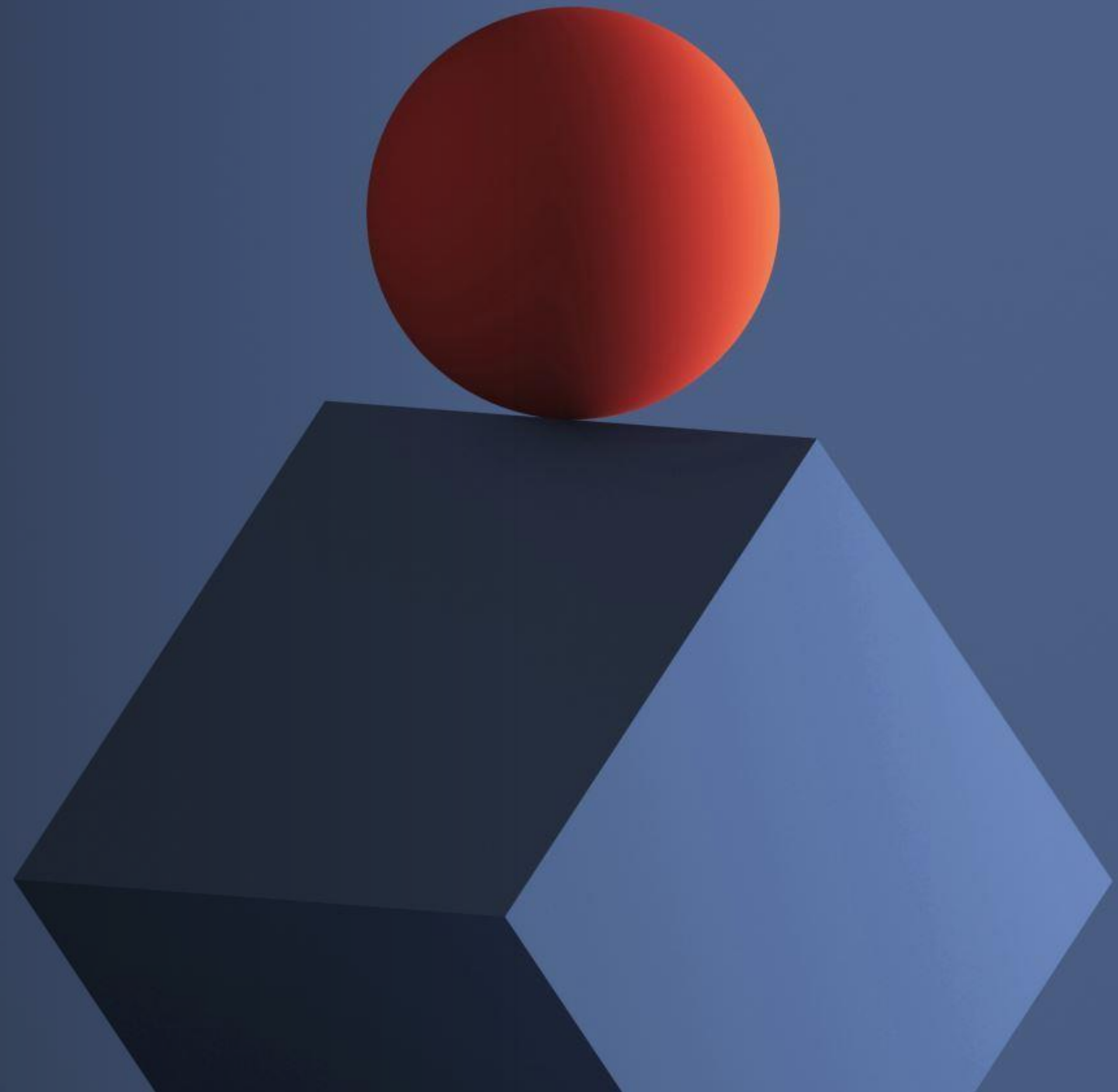
Classification



By Ziad Ahmed

Medical Data

- Assumes features follow a normal (Gaussian) distribution, common in medical data
- Suitable for lab results, imaging intensities, and physiological measurements that show near-normal behavior
- Captures probabilistic relationships between test indicators and disease presence
- Provides interpretable likelihoods rather than black-box outputs
- Enables clinicians to quantify uncertainty and make data-driven diagnostic decisions

Data Exploration

01

The dataset consists of multiple numerical features, analyzed to understand their relationships.

02

Visualized correlations to identify redundancy and dependencies among features.

03

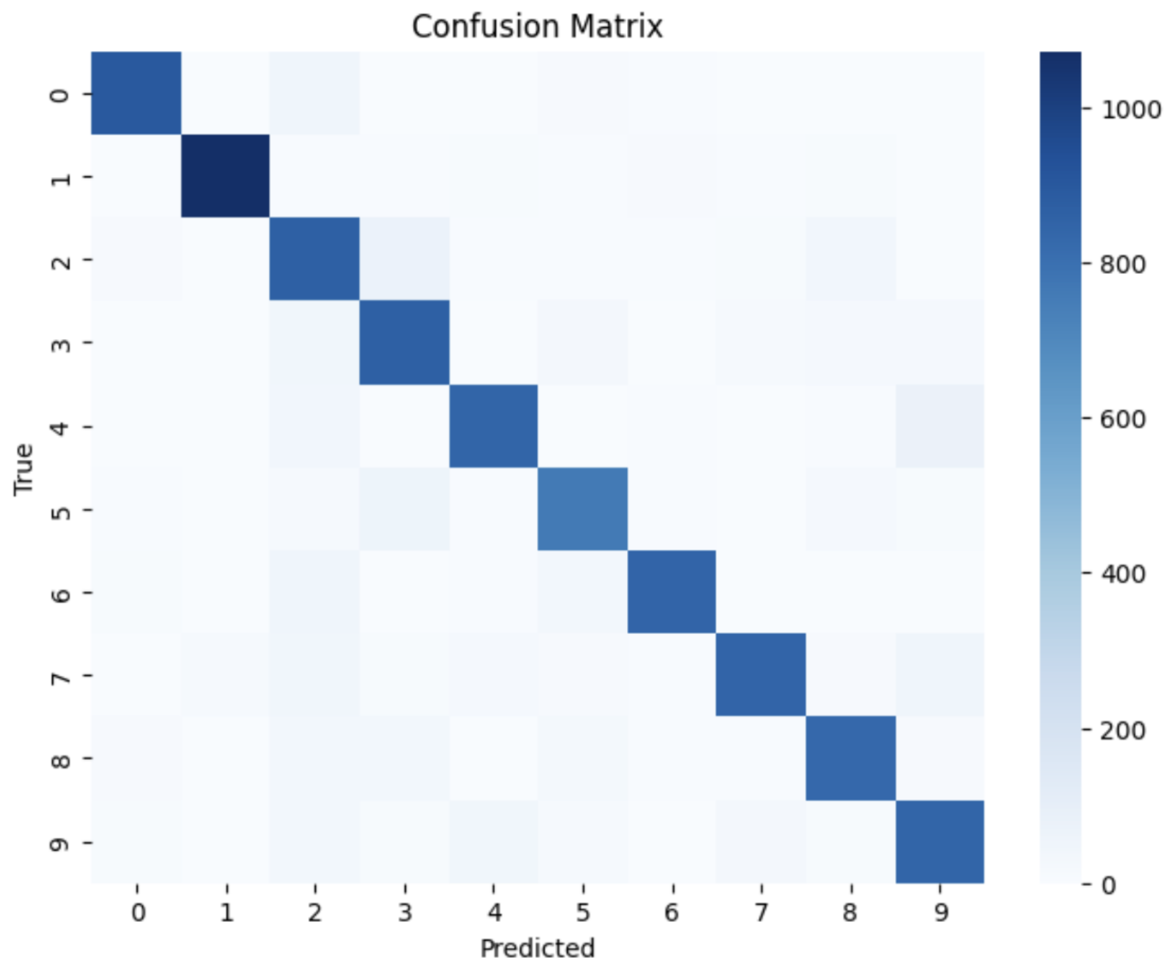
Features with high correlation provide similar information, which may reduce model efficiency.

Model Performance

- The Gaussian classifier achieved 87% test accuracy, demonstrating strong diagnostic capability
- Recall of 0.86 shows the model effectively identifies true positive medical cases
- Precision of 0.85 confirms reliable prediction of disease presence with minimal false alarms
- F1-score of 0.85 indicates a balanced trade-off between sensitivity and specificity
- Low train–test variance reflects stable generalization and minimal overfitting
- Validation through confusion matrix and ROC curve confirms clear class separation and diagnostic reliability

Model Interpretation

- The Gaussian Naive Bayes model estimates how likely each medical test result belongs to each diagnosis class
- Each feature is assumed to follow a normal distribution, defined by its mean and variance for each condition
- For every patient sample, the model calculates the probability density of observed values within each class
- These probabilities are combined using Bayes' theorem to compute the overall likelihood of each diagnosis
- The class with the highest posterior probability is chosen as the model's prediction



Clinical Insights

Model effectively distinguishes normal vs abnormal lab profiles

Misclassifications often occur in borderline readings

Reinforces that the Gaussian model captures the natural variability in patient data, mirroring real-world biological distributions.

Can support early detection before clinical symptoms arise

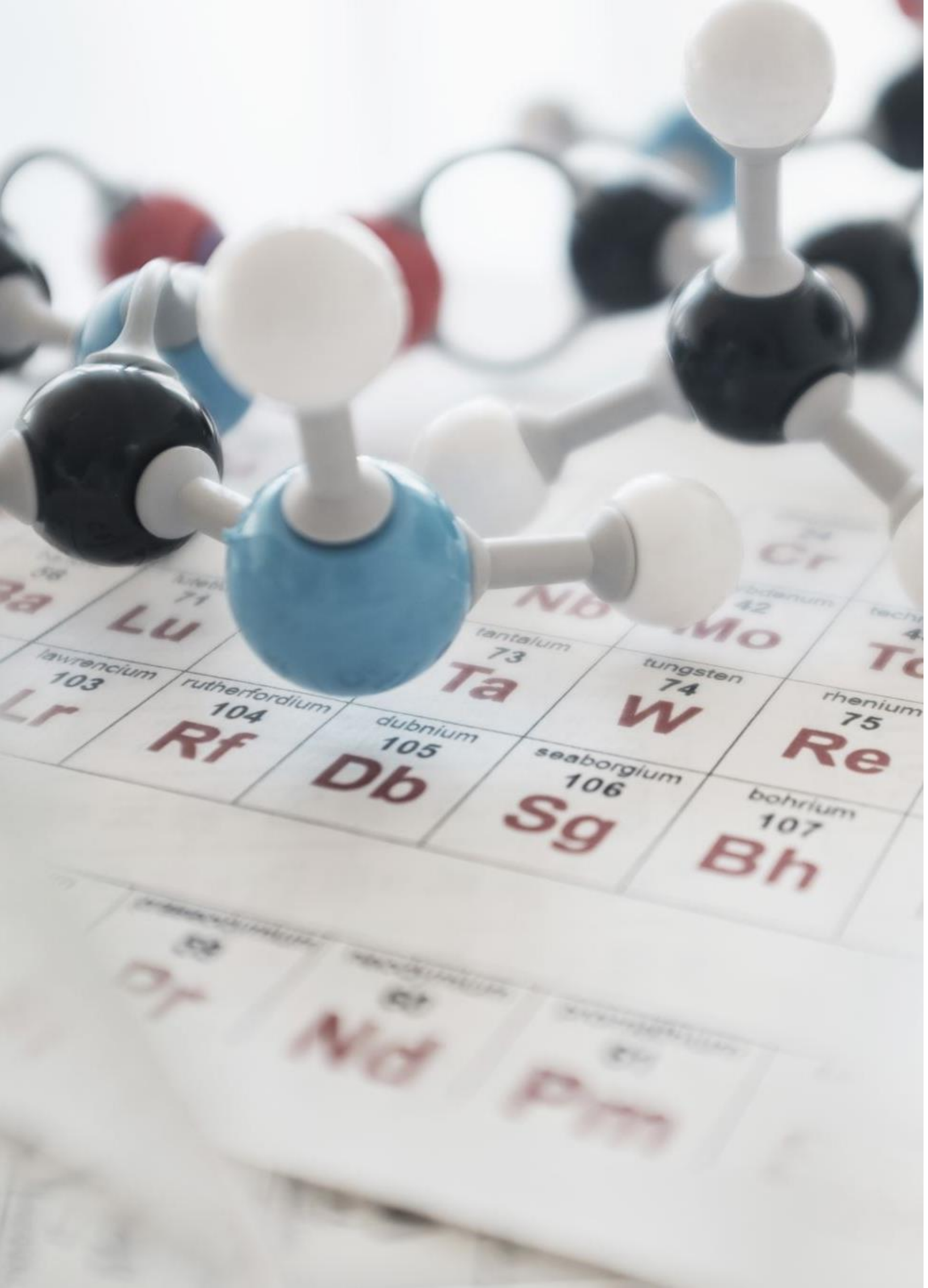
Practical Applications

Diagnostic Support: Suggests potential health risks from standard lab panels

Preventive Screening: Identifies early-warning signs for chronic diseases

Clinical Triage: Flags patients for immediate physician review

Monitoring: Tracks patient improvement or deterioration over time



Conclusion & Future Work

Gaussian Naive Bayes achieved 87% classification accuracy on patient lab data
Assumes each medical feature follows a normal distribution
Demonstrated that probabilistic modeling can effectively classify medical test results with interpretability

- Incorporate additional clinical features (vitals, demographics)
- Compare with models like Random Forest or Logistic Regression
- Integrate probabilistic outputs into a clinical decision-support system

Questions