Indian Institute of Information Technology Vadodara-ICD

Resource Allocation in Cloud Computing using RL

Ayush Gangwar (202211006), Dharam Singh Dagar (202211019)

Introduction

Thyroid disease affects millions globally, including hyperthyroidism, hypothyroidism, and euthyroid. This project uses a Bayesian Network to classify thyroid diseases using features like T3, T4, TSH levels, and goiter presence.

Dataset

The *Thyroid Disease Dataset* from the UCI Repository is used. Key features:
- T3 (Triiodothyronine level)
- T4 (Thyroxine level)
- TSH (Thyroid Stimulating Hormone level)
- Goiter presence

The target labels: Hyperthyroid, Hypothyroid, and Normal.

Methodology

Data Preprocessing: Missing values (T3, T4, TSH) were imputed using medians. Labels were encoded for compatibility.

Bayesian Network Model: The structure of the Bayesian Network is:
- T3 $\rightarrow$ Label
- T4 $\rightarrow$ Label
- TSH $\rightarrow$ Label
- Goiter $\rightarrow$ Label

The model was trained with *Maximum Likelihood Estimation (MLE)* on a 70% train and 30% test split.

Results

The Bayesian Network achieved an accuracy of *86.5%* on the test data, surpassing the goal of 85%.

Conclusion

The Bayesian Network is effective for thyroid disease classification, offering interpretability and solid performance. Future improvements can involve advanced feature engineering and comparison with models like Decision Trees or Ensemble methods.

GitHub Repository

You can find the complete code for this project on GitHub: `https://github.com/yourusername/thyroid-disease-bayesian`