

Implementation of a Computational Rationality Agent for Breast Cancer Diagnosis

Student Name: Luja Shakya
Student Number: A00073470

Foundations of Artificial Intelligence (CMP30L042)

Word Count: [923]

December 7, 2025

Contents

1 Project Overview	2
1.1 Project Aim: The Agent-Environment Framework	2
2 Research Background	2
2.1 Key AI Concepts Used	2
3 Implementation (Environment and Agent Design)	3
3.1 Environment and Data Preprocessing	3
3.2 Validation Strategy	3
4 Experiments, Results and Evaluation	3
4.1 Results Summary	3
4.2 Interpretation and Evaluation	4
5 Conclusions and Discussion	5
5.1 Bounded Rationality	5
5.2 Failure Analysis	6
5.3 Rationality vs. Accuracy	6
Appendices	8
A Gantt Chart	8

1 Project Overview

1.1 Project Aim: The Agent-Environment Framework

This study implements a classical Artificial Intelligence (AI) agent designed to classify breast tumours as Malignant (M) or Benign (B). Building upon the theoretical framework established in Coursework 1, the system is defined as follows:

- **The Agent:** A supervised machine learning classifier (Decision Tree, SVM, or Naïve Bayes) acting as a diagnostic support tool.
- **The Environment:** The Wisconsin Diagnostic Breast Cancer (WDBC) dataset, representing a stochastic clinical setting bounded by 30 specific biophysical features derived from Fine Needle Aspirate (FNA) images.
- **The Objective (Utility Function):** The agent does not seek pure accuracy. Instead, it seeks to maximise Expected Utility in a domain with asymmetric costs. A False Negative (classifying a malignant tumour as benign) has a catastrophic cost (potential patient mortality), whereas a False Positive incurs a lower cost (anxiety/unnecessary biopsy).

Therefore, a "rational" agent in this specific environment is defined not by perfect accuracy, but by maximising Recall (Sensitivity) to ensure patient safety, while maintaining acceptable Precision to reduce healthcare burden.

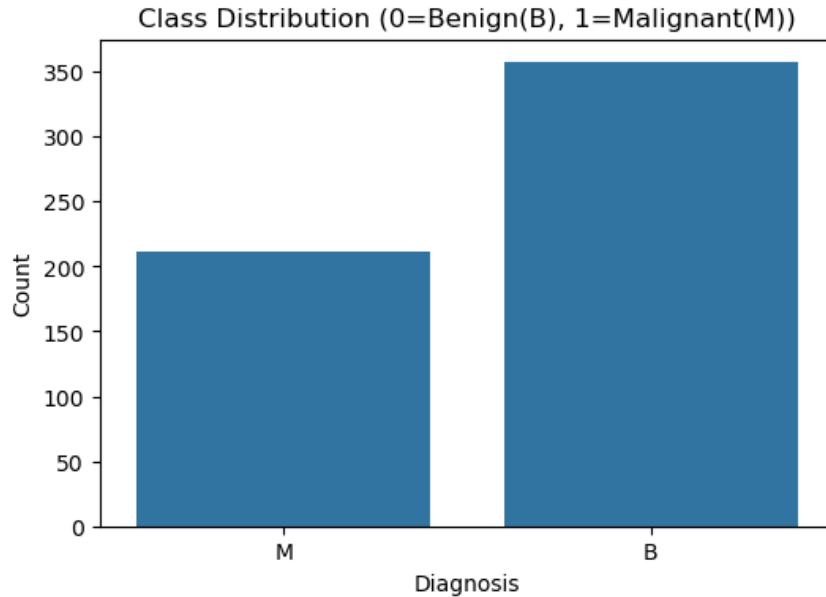


Figure 1: Class Distribution

2 Research Background

2.1 Key AI Concepts Used

Three distinct model architectures were implemented to represent different approaches to rationality within the agent design:

- 1. Decision Tree (Procedural Rationality):** Selected for its "white-box" transparency. It creates hierarchical IF-THEN rules that mimic human diagnostic reasoning. To prevent overfitting (a failure of generalisation), the tree depth was constrained.
- 2. Support Vector Machine (Substantive Rationality):** Selected for its ability to handle high-dimensional data effectively. A Radial Basis Function (RBF) kernel was used to map non-linearly separable data into a higher dimension where a clear margin could be established .
- 3. Gaussian Naïve Bayes (Probabilistic Inference):** Selected as a baseline for computational efficiency. It operates on the strong assumption of feature independence. While biologically unrealistic (e.g., tumour radius and area are correlated), it provides a probabilistic confidence score rather than just a binary classification.

3 Implementation (Environment and Agent Design)

The implementation pipeline was constructed using Python and the scikit-learn library, adhering to industry-standard practices for reproducibility and modularity.

3.1 Environment and Data Preprocessing

The agent's perception is limited to the provided WDBC dataset. The following preprocessing steps were taken to prepare the environment:

- **Label Encoding:** The target variable Diagnosis was encoded ($M = 1$, $B = 0$). This establishes "Malignant" as the positive class, which is crucial for correctly calculating Recall.
- **Feature Scaling:** Standardization ($z = (x - \mu)/\sigma$) was applied to all 30 input features. This is mathematically necessary for the Support Vector Machine (SVM), which calculates distances in high-dimensional space, and beneficial for the convergence of other algorithms [1].

3.2 Validation Strategy

To ensure the agent acts rationally on unseen patients, the system uses Stratified 10-Fold Cross-Validation. This method splits the 569 instances into 10 subsets, preserving the Malignant-to-Benign ratio in each fold. This provides a robust estimate of generalisation performance, minimising the risk that the agent has simply "memorised" the training data.

4 Experiments, Results and Evaluation

4.1 Results Summary

The models were evaluated using Recall (primary utility metric), F1-Score (harmonic mean of precision/recall), and ROC-AUC (discriminative ability). The average performance across 10 folds is summarised below in Table 1.

Table 1: Performance Metrics Comparison (10-Fold CV)

Model	Mean Recall (Sensitivity)	Mean F1-Score	Mean ROC-AUC
Support Vector Machine (SVM)	0.962121	0.966830	0.995637
Decision Tree	0.886580	0.902629	0.916675
Naïve Bayes	0.895671	0.906102	0.985593

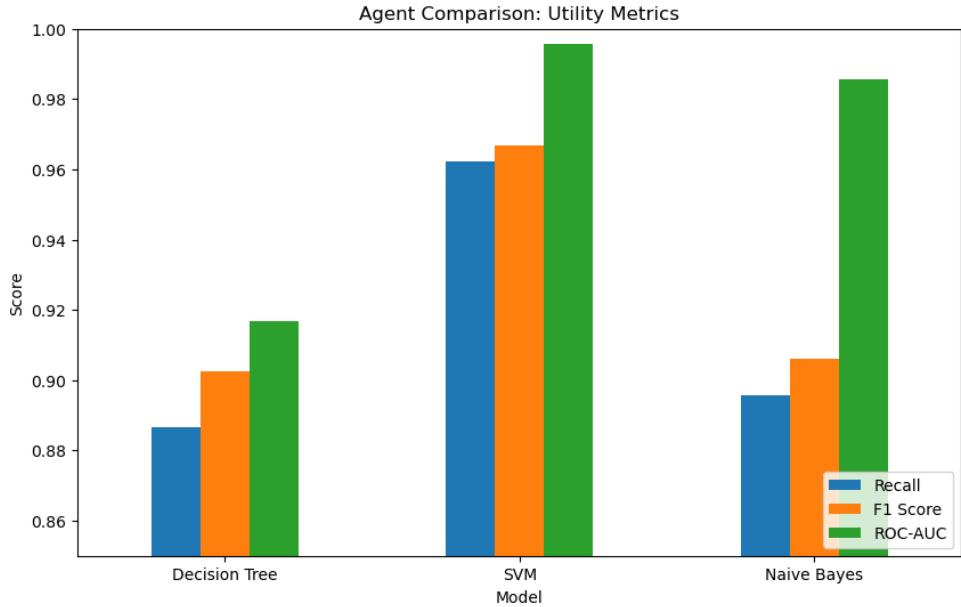


Figure 2: Model Metrics Comparison: Recall, F1 Score, and ROC-AUC. [Source: 81-87]

4.2 Interpretation and Evaluation

Utility Analysis: The SVM demonstrated the highest degree of rationality defined by our utility function. With a Recall of $\approx 96\%$, it minimises the catastrophic error of False Negatives most effectively.

Model Comparison: The Decision Tree, while interpretable, showed higher variance and lower Recall, indicating it is less robust in this environment. Naïve Bayes achieved a high ROC-AUC, suggesting strong discriminative power, but its default decision boundary yielded lower Recall than the SVM.

Visualisations generated in the notebook (Confusion Matrices and ROC Curves) further confirmed that the SVM maintained the highest True Positive rate across various decision thresholds.

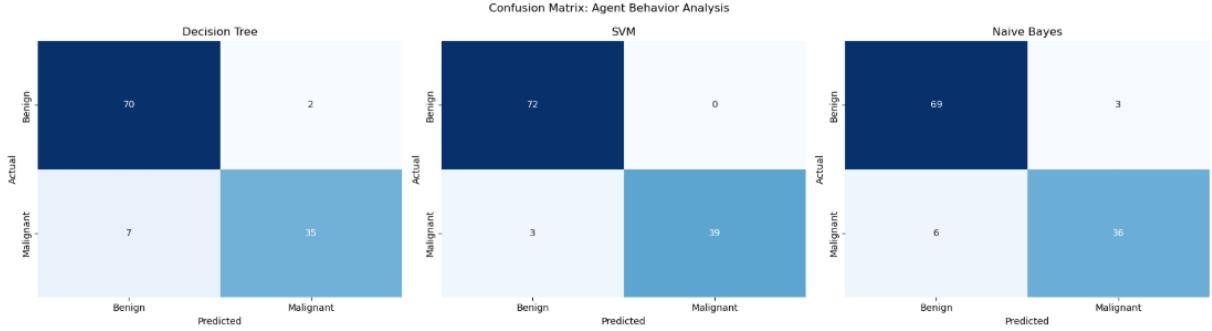


Figure 3: Confusion Matrix Comparison

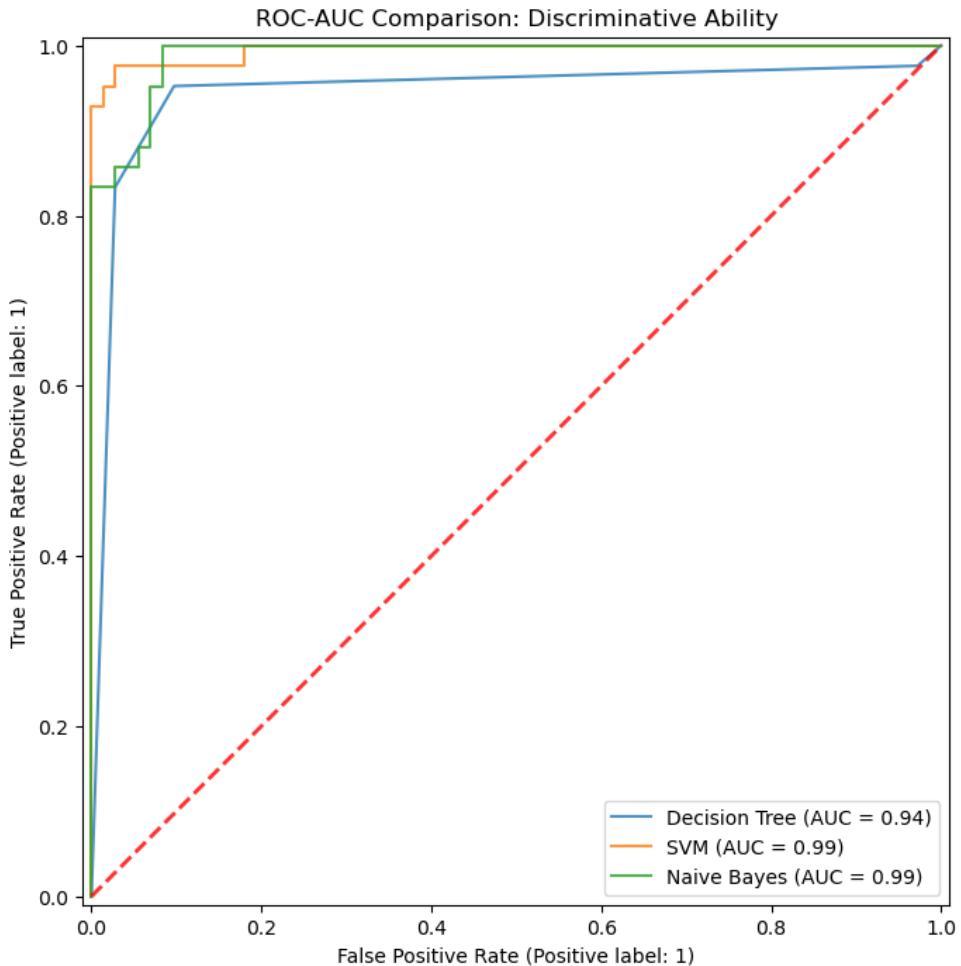


Figure 4: ROC-AUC Curve Comparison (Test Set Evaluation). SVM and Naive Bayes show superior AUC (0.99).

5 Conclusions and Discussion

5.1 Bounded Rationality

This assessment highlights the concept of Bounded Rationality. The Decision Tree agent is bounded by its need for axis-parallel splits, making it incapable of capturing complex

biological non-linearities. In contrast, the SVM exhibits high substantive rationality (excellent outcomes) but lacks procedural rationality. It acts as a "black box," offering no explanation for its diagnosis. This presents a trade-off: we sacrifice the ability to understand the agent's reasoning (interpretability) to achieve the objective of saving more lives (higher Recall).

5.2 Failure Analysis

The Naïve Bayes model underperformed the SVM on Recall. This is likely due to its core assumption of feature independence. In the WDBC dataset, features like `radius_mean`, `perimeter_mean`, and `area_mean` are mathematically coupled. The agent's inability to perceive these correlations represents a cognitive limitation, leading to overconfident but occasionally incorrect probability estimates.

5.3 Rationality vs. Accuracy

As argued in Coursework 1, maximising "Accuracy" would be irrational. An agent that classifies every tumour as Benign might achieve $\approx 63\%$ accuracy (based on class distribution) but would kill every patient with cancer. The implemented pipeline correctly prioritised Recall. However, the SVM still produced a small number of False Positives. In a clinical environment, this is an acceptable "cost" of rationality. It is rational to biopsy a benign tumour to ensure a malignant one is not missed.

References

- [1] [Scikit-learn developers, "Support Vector Machines," Scikit-learn.org. [Online]. Available: <https://scikit-learn.org/stable/modules/svm.html>.]

Appendices

A Gantt Chart

