

Mini Project: Probabilistic Reasoning on System Data

AI/ML Internship Mini Project

Name: Monisha N

Degree: BTech IT

Year: 1st Year

This mini project was completed as part of my **AI/ML internship**. The purpose of this project is to learn how simple machine learning models can be used to understand system behavior and make decisions when data is uncertain. The focus is on understanding and explanation, not on high accuracy.

Only the **easy_queue_data.csv** dataset is used in this project.

Dataset Description

The dataset contains the following columns:

- wait_time: how long a request waited before service
- service_time: how long the request took to be served
- label: 1 means problematic, 0 means normal (these labels may not always be correct)

1. Data Understanding

Most useful feature:

Wait time is the most useful feature because requests that wait longer usually indicate delays or congestion in the system.

Noisy feature:

Service time is more noisy because even normal requests can sometimes take longer to complete.

What is a problematic request?

A problematic request usually has a long wait time and may also have a longer service time.

2. Naive Bayes Model

Naive Bayes is used to estimate the probability that a request is problematic. It does not give a sure answer, but a likelihood based on past data.

Confident and correct example:

In one case, the model gives a high probability for a request that is labeled as problematic. This happens because the wait time is much higher than most other requests.

Confident but wrong example:

In another case, the model gives a high probability for a request that is labeled as normal. This shows that the model can be confident but still make mistakes due to noisy labels or limited features.

3. K-Means Clustering

K-Means is used to group similar requests together without using labels. Two clusters are chosen to represent simple system behavior patterns.

Cluster 0:

Requests with shorter wait times and shorter service times. These represent normal system behavior.

Cluster 1:

Requests with longer wait times and longer service times. These represent delayed system behavior.

4. Reflection

Both models work well when the wait time is clearly very high or very low. They struggle with cases that fall in between. This happens because the models are simple, the amount of information is limited, and the labels are not always correct.

Conclusion

This mini project shows how basic machine learning models can help reason under uncertainty while also highlighting that mistakes are unavoidable in real systems.