Capstone 2: Project Proposal

I would like to do the Kaggle project on breast cancer detection (link: https://www.kaggle.com/competitions/rsna-breast-cancer-detection).

Problem statement formation: How can the RSNA (Radiological Society of North America) develop new data-based methods to evaluate mammograms, to bring down the false positive rate to below 25% within a 5-year period ?

Context: At the present time, early detection of breast cancer requires the expertise of a health specialist. It would be better to develop machine learning techniques to do this, in order to save costs. Furthermore, there is currently a false positive rate, which leads to even more costly procedures. We hope that a machine learning-based method will help bring down the false positive rate.

Criteria for success: The false positive rate is below 25% within 5 years.

Scope of solution space: Classify radiographic breast images from tens of thousands of patients as cancerous or benign. Also, we will use only traditional algorithms, and not neural networks.

Constraints: The medical community might resist the idea of diagnosing cancer by computers, and they might insist that it's important to involve human expertise.

Stakeholders: Radiologic subspecialties from all the countries around the world.

Data sources: Kaggle.