Data Mining in Brain Tumor Detections

By: Tariere Timitimi, Ana Sytina, Tahmidul Haque

PROJECT OVERVIEW STATEMENT	**Project Name:** Brain Tumor Detection Using Data Mining Techniques	**Team Name:** Cancer Cure Miners

Problem/Opportunity:

Early identification of cancer and other life-altering diseases is crucial for enhancing the quality of life and improving patient outcomes. Like other types of cancer, brain tumors have the ability to spread throughout other parts of the body such as more parts of the brain and the spinal cord. This increases the complexity of treatment and isn't a favorable outcome. Key Factors of Brain Tumors are its Timing and its unpredictability. Just like many others, this form of cancer can start off in the area but then will progress to spread the diseases throughout the whole body at a fast pace. That is why, timing is crucial for early detection of possible cancer cells and thereby finding possible and fast treatment plans. When it comes to brain tumor diagnosis, early detection can open possibilities to better treatment options and reduce the risk of the tumor progressing and impacting the patient's life. Using MRI brain scan images to detect brain tumors in patients through the application of data mining techniques we can develop a model capable of accurately detecting brain tumors, which can lead to the reduction of human error and enable a faster method of diagnosis using medical imaging.

Goal:

The aim of our research project is to develop a predictive model using data mining techniques to identify indicators of brain tumors from MRI images. By analyzing a dataset of brain tumor images and applying deep learning algorithms, we hope to improve the accuracy and efficiency of tumor detection which facilitates timely diagnosis and treatment. We will follow a structured approach, with each team member handling phases that will be determined through-out the semester. Using publicly available datasets from Kaggle and established techniques, we ensure feasibility while maintaining a clear timeline for development, testing, and evaluation. The model's performance will be assessed through key metrics such as accuracy, precision, recall, and F1-score to measure effectiveness.

Objectives:

- Data Collection & Preprocessing (Weeks 1–3)
  - Outcome: Gather and look over our chosen dataset of brain MRI images.
  - Measure: A well-structured dataset ready for analysis.
  - Action: Handle missing values and normalize features.
- Exploratory Data Analysis (Weeks 3–5)
  - Outcome: Identify key trends and correlations.
  - Measure: Summary statistics and visual insights.
  - Action: Incorporating various data mining tools for analysis.
- Model Development & Training (Weeks 5–9)
  - Outcome: Implement predictive models (Decision Trees, SVM, Neural Networks).

    - Measure: Accuracy and efficiency metrics.
    - Action: Train, validate, and fine-tune models.
- Model Evaluation & Optimization (Weeks 9–11)
    - Outcome: Assess and refine model performance.
    - Measure: Metrics like precision, recall, and F1-score.
    - Action: Compare models and optimize parameters.
- Final Report & Presentation (Weeks 11–12)
    - Outcome: Summarize findings and results.
    - Measure: Comprehensive report.
    - Action: Document methodology and conclusions.

Success Criteria:

The project will be considered successful if:

- The predictive model achieves an accuracy of at least **85%** in detecting brain tumors from the dataset
- Key features influencing breast cancer detection are identified and validated.

Assumptions, Risks, Obstacles:

This project assumes that the chosen dataset is both comprehensive and representative of real-world brain tumor cases and that the chosen machine learning models are well-suited for the task. There are several challenges that may impact performance, such as data quality issues and time constraints. Additionally, access to high-quality, diverse datasets can be limited, and interpreting complex models like deep learning algorithms can make validation more challenging. Overcoming these obstacles is a goal that is important to ensuring the model's reliability and effectiveness in medical applications.