WQD 7005 – 2024/2025 – S2

Project (Due: Week 13)

Objective:

Leverage AI technologies, including Large Language Models (LLMs), Small Language Models (SLMs), and Generative AI (GenAI), to engineer features and build predictive models for anticipating patient health deterioration based on collected vital signs and health questionnaire responses.

Tasks:

1. **Dataset Simulation and Feature Engineering (4 marks)**
    - Utilize GenAI to generate realistic patient data, including vital signs and detailed textual questionnaire responses.
    - Apply LLMs to extract meaningful features from simulated textual data (clinical notes, health records).
2. **Predictive Model Development (5 marks)**
    - Construct and evaluate predictive models, including traditional models (Random Forest, XGBoost, Neural Networks) and advanced Transformer-based models.
    - Use SLMs for specialized NLP tasks like sentiment analysis, clinical text interpretation, and classification of questionnaire responses.
3. **Model Evaluation and Interpretation (4 marks)**
    - Evaluate model performance using metrics such as accuracy, F1-score, ROC-AUC, etc.
    - Leverage LLMs to summarize and interpret complex model outputs, explaining performance clearly and concisely.
4. **Comprehensive AI-Assisted Final Report (2 marks)**
    - Provide a detailed report (5-7 pages) clearly outlining the methodology, analytical techniques, results comparison, conclusions, and future recommendations.
    - Explicitly disclose the use of AI tools within your report.

Deliverables:

- Comprehensive Jupyter notebook with documented code, results, and AI-assisted interpretations.
- Final report clearly outlining methodology, analyses, findings, model comparisons, conclusions, and recommendations.

AI Usage Guidelines:

- Clearly disclose all LLM, SLM, and GenAI usage.
- AI-generated assistance is allowed, but substantial direct AI-generated content is prohibited.
- Random oral assessments may be conducted to ensure authenticity of submitted work.