Unit 3: Methodology and Research Methods

e-Portfolio Activity: Research Proposal Review

Task:

Considering your thoughts on your chosen area of interest for your project:

- Which of the methods described in this week's reading would you think would suit your purpose?
- Which data collection methods would you consider using?
- Which required skills will you need to have or develop for the chosen project?

Answers:

1. Which of the methods described in this week's reading would you think would suit your purpose?

Mixed methods involving qualitative and quantitative approaches. The use of mixed methods, which integrate both qualitative and quantitative research approaches, is particularly well suited for this study on implementing ML tools in diabetes diagnosis. This methodology provides a comprehensive understanding by combining numerical insights and experiential perspectives. The complexity of ML models in healthcare, including issues related to data pre-processing, algorithm selection, and user acceptance, necessitates diverse forms of data collection and analysis. Quantitative methods are essential for evaluating the performance of ML models and analysing structured data from healthcare records. Conversely, qualitative approaches such as interviews with healthcare professionals and data scientists provide in-depth insights

into practical implementation challenges and ethical concerns. Combining these perspectives allows for a holistic understanding of the technical and contextual factors that influence the adoption of ML solutions in diabetes diagnostics. Moreover, mixed methods research aligns with the pragmatic paradigm, which emphasizes the use of diverse approaches to address research questions effectively. This paradigm ensures that the study remains adaptable and focused on real-world applications and outcomes. By leveraging both data types, this methodology facilitates triangulation of findings, enhancing the validity and reliability of the research. This ensures that the conclusions drawn are evidence-based and context-sensitive, making the results valuable for healthcare and data science stakeholders.

2. Which data collection methods would you consider using?

Data collection is critical for building an effective ML model. For diabetes diagnosis, the following data collection methods could be considered:

- Public Datasets: Use existing datasets like the Pima Indians Diabetes Dataset (available on platforms like Kaggle or UCI Machine Learning Repository), which includes features such as glucose levels, BMI, age, and insulin levels.
- Electronic Health Records (EHRs): Collaborate with healthcare providers to access anonymized patient data, including lab results, medical history, and demographic information.
- Wearable Devices: Collect real-time data from wearable devices that monitor glucose levels, physical activity, and other health metrics.
- Surveys and Questionnaires: Gather additional data on lifestyle factors (e.g., diet, exercise, smoking habits) that may influence diabetes risk.
- Medical Imaging: If applicable, collect data from medical imaging techniques (e.g., retinal scans for diabetic retinopathy).

- Data Preprocessing: Ensure data is cleaned, normalized, and free from missing values or outliers. Techniques like imputation, feature scaling, and encoding categorical variables will be necessary.

3. Which required skills will you need to have or develop for the chosen project?

To successfully implement this project, the following skills are essential:

- **Programming Skills:**
  - Proficiency in Python or R, as they are widely used for ML.
  - Familiarity with libraries like Scikit-learn, TensorFlow, Keras, Pandas, NumPy, and Matplotlib/Seaborn for data manipulation, modeling, and visualization.
- **Machine Learning Knowledge:**
  - Understanding of supervised and unsupervised learning algorithms.
  - Knowledge of model evaluation metrics (e.g., accuracy, precision, recall, F1-score, ROC-AUC).
- **Data Handling and Preprocessing:**
  - Ability to clean, preprocess, and analyze datasets.
  - Experience with handling missing data, feature engineering, and dimensionality reduction (e.g., PCA).
- **Statistical Analysis:**
  - Understanding of statistical concepts to interpret data and model results.
- **Domain Knowledge:**
  - Basic understanding of diabetes, its risk factors, and diagnostic criteria.
- **Model Deployment:**
  - Familiarity with deploying ML models using frameworks like Flask, FastAPI, or cloud platforms (e.g., AWS, Google Cloud).

- **Soft Skills:**
    - Problem-solving, critical thinking, and the ability to communicate findings effectively to non-technical stakeholders.