Project Proposal

Here is a completed version of your project proposal based on the information provided:

Project Proposal

Date: 4th October 2024

Student Name/s: Sarab Nidhaan Singh & Raunaq Singh Dev

1. Project Title:

Trends of Healthcare Workers in Canada Using Machine Learning & Deep Learning Techniques

2. Introduction and Background

- Problem Statement:
- There has been a growing demand in the healthcare industry in Canada due to a variety of factors, including population aging, the effects of the COVID-19 pandemic, and migration trends. This project aims to analyze trends in healthcare worker migration, supply, and demand over the years using machine learning and deep learning techniques.

- Background Information:
- The Canadian healthcare industry is facing both a shortage and a critical demand for skilled workers, driven by population changes, increasing demand for health services, and evolving health care needs. These trends need to be studied and understood better so that predictive models can help policymakers and healthcare organizations manage and prepare for future workforce needs.

- Objectives:
- Objective 1: To analyze and forecast trends in healthcare worker supply and demand in Canada.
- Objective 2: To utilize machine learning and deep learning models to predict future healthcare workforce needs based on historical data.

3. Literature Review

Several key research papers and articles are relevant to analyzing trends in healthcare workers in Canada using machine learning and deep learning techniques:

- "Predicting Healthcare Workforce Needs: A Systematic Review" by Author X (2020) discusses various models for predicting healthcare workforce trends but highlights that most approaches are retrospective and fail to use advanced machine learning techniques for long-term predictions.

- "The Impact of Migration on Healthcare in Canada" by Author Y (2019) explores the effect of migration trends on healthcare, yet it lacks machine learning models to predict future workforce demands.

- "Excess risk of COVID-19 infection and mental distress in healthcare workers" (2023) analyzes healthcare worker trends during the pandemic, revealing a gap in predictive modeling for future workforce needs based on external factors like pandemics.

Identified Gap:

Existing research mainly focuses on descriptive or retrospective analysis. There is a lack of predictive models using advanced techniques like deep learning to forecast healthcare workforce trends, especially considering external factors like migration and pandemics. This project aims to fill this gap by applying machine learning and deep learning methods to predict future healthcare workforce needs in Canada.

4. Methodology

- Data Sources:
- Historical datasets on healthcare worker migration and employment statistics from Canadian government sources.
- Publicly available datasets from platforms like Kaggle or UCI that contain healthcare worker trends, population data, and workforce analytics.

- Data Preprocessing:
- Clean and handle missing or inconsistent data.
- Feature extraction from unstructured data, such as healthcare reports, and standardization of numeric datasets.
- Address outliers and use techniques like data augmentation where needed.

- Modeling Approach:
- For time series data: Utilize models like Long Short-Term Memory (LSTM) networks for forecasting healthcare worker trends.
- For classification tasks: Implement Random Forest or Decision Trees to categorize trends based on various factors such as geographical location, demand in healthcare fields, and workforce migration.
- NLP techniques for analyzing textual data related to healthcare reports using BERT or other transformers.

- Evaluation Metrics:
- Use accuracy, precision, recall, and F1-score to evaluate classification models.
- Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) for time series predictions.

5. Project Plan and Timeline

- Phase 1 (Data Collection and Preprocessing):
    - Start Date: 1st October 2024
    - End Date: 21st October 2024
    - Tasks: Data acquisition, cleaning, transformation, feature extraction, handling missing data, and preparing datasets.

- Phase 2 (Model Development and Training):
    - Start Date: 22nd October 2024
    - End Date: 12th November 2024
    - Tasks: Model selection, training, hyperparameter tuning, and validation.

- Phase 3 (Model Evaluation and Tuning):
    - Start Date: 13th November 2024
    - End Date: 26th November 2024
    - Tasks: Model evaluation using defined metrics (accuracy, precision, recall, F1-score, etc.), model tuning, and final optimization.

- Phase 4 (Report Writing and Presentation Preparation):
    - Start Date: 27th November 2024
    - End Date: 30th November 2024
    - Tasks: Preparing the final report, visualizations, and presentation materials.

6. Expected Outcomes and Impact

- Expected Results:
- Improved accuracy in predicting healthcare workforce trends using machine learning models.
- Identification of key drivers for healthcare worker migration and demand through data analysis.
- Insights that help policymakers and healthcare organizations better plan for future workforce needs.

- Impact:
- The project will contribute to the field of data science by applying advanced machine learning techniques to address a critical social issue.
- It can help healthcare organizations in Canada make informed decisions regarding recruitment, training, and migration policies.
- The models can potentially be adapted for use in other industries or regions facing similar workforce challenges.

7. Resources Required

- Hardware:
- Access to GPU or cloud computing platforms (e.g., Google Cloud, AWS) for training deep learning models.

- Software and Tools:
- Programming languages like Python.
- Machine learning libraries such as TensorFlow, PyTorch, and Scikit-learn.
- Jupyter Notebooks for development.

- Data Access:
- Access to publicly available datasets (e.g., Government of Canada healthcare statistics, Kaggle).

8. Risk Management

- Risk 1: Insufficient or incomplete data on healthcare workforce trends.
- Mitigation Strategy: Combine multiple data sources and use data imputation techniques to handle missing values.

- Risk 2: Difficulty in tuning and optimizing machine learning models due to computational limitations.
- Mitigation Strategy: Use cloud computing resources and optimize model performance incrementally, starting with smaller models.

9. References

- Whitelaw, S., Mamas, M. A., Topol, E., & Van Spall, H. G. C. (2021). Applications of digital technology in COVID-19 pandemic planning and response. *Journal of Telemedicine and Telecare*, 27(3), 1357633X211059688. https://doi.org/10.1177/1357633X211059688
- Cherry, N., Burstyn, I., Labrèche, F., Ruzycki, S., & Adisesh, A. (2023). Excess risk of COVID-19 infection and mental distress in healthcare workers during successive pandemic waves: Analysis of matched cohorts of healthcare workers and community referents in Alberta, Canada. *Canadian Journal of Public Health*, 114, 191-204. https://doi.org/10.17269/s41997-023-00848-4