Summary of Collaborative Discussion 1

why AI is now ubiquitous and why it is important for companies to invest in AI technologies

Artificial Intelligence (AI) is a technology where computer applications can learn and acquire knowledge from vast set of data and can perform intellectual tasks similar to human beings. In recent times, AI has been playing an important role in economic, social, healthcare and scientific areas of businesses by helping in decision making process, relieving human workforce from repetitive work environment, increasing production of goods and services and accelerating scientific research. While AI provides enormous benefits, it is important for companies to understand challenges in data governance.

In our discussion about AI utilisation in healthapps, the aids such as initial contact with health worker, monitoring patient symptoms and reduced human involvement around the clock for each patient were discussed. Many of my peers had agreed with the benefits of healthapps and also pointed out the importance of data management and the consequences, if breached. Considerations such as patient safety, data protection, setting principles and ethical standards, accuracy of results and potential risks involved using AI were discussed referring to the guidelines of the Alan Turing Institute (Leslie, 2019) and the WHO (World Health Organisation, 2021).

World Health Organisation (2021) had laid down core principles to ensure safety, transparency, responsibility, equality in healthcare applications and to implement these principles it is essential that all stakeholders, including patients, healthcare providers, AI application developers work together to integrate ethical standards at every phase of application design, development, and deployment.

Based on the available governance framework, the companies when developing AI applications have to include these principles on ethics, safety and quality in their organisational objectives, policies, processes and procedures, proactively include impact risk assessment, policy for continuous quality improvement and learning from near-miss and errors from wider community. In addition companies can enhance its public confidence by displaying its quality compliance with international organisation for standardisation.

Reference:

Leslie, D. (2019) Understanding artificial intelligence ethics and safety: A guide for the responsible design and implementation of AI systems in the public sector. The Alan Turing Institute.

World Health Organization. (2021) Ethics and Governance of Artificial Intelligence for Health. Available from: https://www.who.int/publications/i/item/9789240029200 [Accessed 27 Mar 2022].

Summary of Collaborative Discussion 2

The discussion was about the two machine learning algorithms, decision tree and clustering.

Decision tree algorithm is an example of supervised learning algorithm, simple to understand, deals with classification and regression problems and used with categorical, binary and numerical data. Dewaji (2022) had provided comprehensive description about this algorithm with example and method of tree construction with calculation of entropy and information gain. ID3 and C4.5 are the popular algorithms with decision tree and Dada (2022) had discussed about the efficiency of these algorithms.

Clustering is an unsupervised machine-learning algorithm commonly used to identify and group similar data from large dataset. Moodley (2022) had broadly discussed about the k-means clustering algorithm and methods to calculate the k value.

 Every learning algorithm has certain strengths and weaknesses, it is important for the data scientist to decide the right algorithm for the given research problem based on the input data available and the desired output requirement. In addition to the performance accuracy, the algorithm should have good interpretability, performance dependency on number of data points and features, data format it can handle, linearity of data, optimal training and prediction time and memory requirements in the server or system.

Data mining methodology, CRISP-DM (cross-industry standard process), helps in following a systematic process, consisting of phases such as business and data understanding, data preparation and modelling, evaluation and deployment.

Feature selection is an important step in feature engineering process that helps in selecting a small set of features from a large pool of features either manually or automatically, thereby reducing the computation time and improving quality of results (Bell, 2020). For data with thousands of features, dimensionality reduction algorithms help to find highly significant features with linear methods such as linear discriminant analysis, principal component analysis and non-linear methods such as locally linear embedding, kernel learning and neural networks (Balogun, 2020)

For large data, Gradient boosting algorithms such as GBM, XGBoost, LightGBM provide the goodness of using ensemble of learning algorithms to increase the prediction power of the model. Ganapavarapu (2022) had pointed out the advantages with ensemble learning in better predictions than the other routine machine-learning algorithms.

References:

Balogun, A.O., et al. (2020) Impact of feature selection methods on the predictive performance of software defect prediction models: An extensive empirical study. Symmetry 12(7): 1147.

Bell, J (2020) Machine Learning: Hands-On for Developers and Technical Professionals. 2nd Ed. Wiley. John Wiley & Sons P&T.

Dada, M. (2022) Peer response to discussion, Initial post: Decision tree and clustering algorithms, 23 April.

Dewaji, I. (2022) Peer response to discussion, Initial post: Decision tree and clustering algorithms, 20 April.

Ganapavarapu, L. (2022) Peer response to discussion, Initial post: Decision tree and clustering algorithms, 24 April.

Moodley, K. (2022) Peer response to discussion, Initial post: Decision tree and clustering algorithms, 21 April.