Graduate Certificate in Adopting AI for Infection Prevention and Control

Machine Learning and Predictive Modeling

Artificial Intelligence (AI): The simulation of human intelligence processes by machines, especially computer systems. These processes include learning (the acquisition of information and rules for using the information), reasoning (using the rules to reach approximate or definite conclusions), and self-correction.

Deep Learning: A subset of machine learning that is based on artificial neural networks with representation learning. It can process a wide range of data resources, is not dependent on feature engineering, and can learn and improve from experience.

Gradient Boosting: A machine learning technique used for regression and classification tasks. It builds an additive model in a forward stage-wise fashion, allowing for the optimization of arbitrary differentiable loss functions.

Infection Prevention and Control (IPC): A practical, evidence-based approach that is used to prevent and control the spread of infections in healthcare settings. IPC involves the implementation of systems and strategies that reduce the risk of infection for patients, healthcare workers, and visitors.

Logistic Regression: A statistical model that is used for binary classification tasks. It estimates the probability of an event occurring, based on prior knowledge of one or more predictor variables.

Machine Learning (ML): A type of artificial intelligence that allows systems to learn and improve from experience without being explicitly programmed. It involves the use of algorithms to analyze and draw insights from data.

Natural Language Processing (NLP): A field of artificial intelligence that focuses on the interaction between computers and humans through natural language. The ultimate objective of NLP is to read, decipher, understand, and make sense of human language in a valuable way.

Neural Networks: A type of machine learning model that is inspired by the human brain. It is composed of interconnected layers of nodes, or "neurons," and is used for a variety of tasks, including image and speech recognition, natural language processing, and predictive modeling.

Overfitting: A modeling error that occurs when a model is excessively complex, such as when it has too many parameters relative to the number of observations. Overfitting can result in a model that performs well on the training data but poorly on new, unseen data.

Predictive Modeling: The process of creating a mathematical model that can predict an outcome based on a set of input variables. Predictive modeling is used in a variety of fields, including finance, healthcare, and

marketing, to make informed decisions and predictions.

Random Forests: A machine learning technique used for classification and regression tasks. It builds multiple decision trees and combines their outputs to improve the accuracy and stability of the model.

Reinforcement Learning: A type of machine learning in which an agent learns to make decisions by interacting with its environment. The agent is rewarded or penalized based on the quality of its decisions, and it uses this feedback to improve its performance over time.

Regression: A statistical method used for predicting a continuous outcome variable based on one or more predictor variables. There are several types of regression, including linear regression, logistic regression, and polynomial regression.

Support Vector Machines (SVMs): A type of supervised machine learning algorithm that can be used for classification or regression tasks. SVMs work by finding the line, or "hyperplane," that best separates the data into different classes.

Supervised Learning: A type of machine learning in which the model is trained on a labeled dataset, meaning that the input variables and the corresponding output variable are known. The model uses this information to learn the relationship between the input and output variables and to make predictions on new, unseen data.

Unsupervised Learning: A type of machine learning in which the model is trained on an unlabeled dataset, meaning that the input variables are known but the output variable is not. The model must find patterns and relationships in the data on its own, without any prior knowledge of the desired outcome.

Validation Curve: A graphical tool used to diagnose issues with model performance, such as overfitting or underfitting. It shows the relationship between the model's performance and the complexity of the model, and can help identify the optimal level of complexity for the model.

Visualization: The process of representing data in a visual format, such as a chart, graph, or map. Visualization can help make complex data more understandable and accessible, and can reveal patterns and trends that might otherwise go unnoticed.