
Professional Certificate in AI in Public Health and Safety

Machine Learning Techniques for Health Surveillance

Association rule mining: A machine learning technique used to discover interesting relationships between variables in large datasets. It involves identifying patterns where one event leads to another in a probabilistic manner. For example, association rule mining can be used to find relationships between symptoms and diseases in health surveillance data.

Artificial neural networks (ANNs): A type of machine learning model inspired by the structure and function of the human brain. ANNs consist of interconnected nodes that process information and learn patterns through training data. They are commonly used in health surveillance for tasks such as image recognition and predictive modeling.

Big data: Large and complex datasets that cannot be easily processed using traditional data processing applications. Big data is characterized by volume, velocity, and variety. In health surveillance, big data may include electronic health records, medical imaging data, and genomics data.

Classification: A machine learning technique used to categorize data into predefined classes or labels based on input features. Classification algorithms are commonly used in health surveillance for tasks such as disease diagnosis and risk prediction.

Clustering: A machine learning technique used to group similar data points together based on their characteristics. Clustering algorithms are useful for identifying patterns and structures in data without predefined labels. In health surveillance, clustering can be used to segment populations based on health status or risk factors.

Deep learning: A subset of machine learning that utilizes artificial neural networks with multiple layers to learn complex patterns from data. Deep learning models can automatically discover features from raw data and are capable of handling large-scale datasets. In health surveillance, deep learning is used for tasks such as image analysis and natural language processing.

Ensemble learning: A machine learning technique that combines multiple models to improve predictive performance. Ensemble methods such as random forests and gradient boosting are commonly used in health surveillance to reduce overfitting and increase accuracy.

Feature engineering: The process of selecting, transforming, and creating new features from raw data to improve the performance of machine learning models. Feature engineering is crucial in health surveillance to extract relevant information from diverse datasets and optimize model outcomes.

Health surveillance: The ongoing systematic collection, analysis, interpretation, and dissemination of health-

related data to inform public health decision-making. Machine learning techniques are increasingly being applied in health surveillance to automate data processing, identify trends, and predict health outcomes.

Imbalanced data: Datasets in which the distribution of classes or labels is skewed, with one class significantly outnumbering the others. Imbalanced data is common in health surveillance, where rare events such as disease outbreaks or adverse drug reactions need to be accurately detected using machine learning models.

Interpretable machine learning: Machine learning models that are designed to provide transparent and understandable explanations for their predictions. Interpretable machine learning is essential in health surveillance to ensure that decisions based on AI algorithms are trustworthy and actionable.

K-means clustering: A popular clustering algorithm that partitions data points into k clusters based on their similarity. K-means clustering is used in health surveillance to group patients with similar characteristics or behaviors for targeted interventions and personalized healthcare.

Logistic regression: A statistical model used for binary classification tasks, where the outcome variable has two possible outcomes. Logistic regression is commonly used in health surveillance to predict the probability of disease occurrence or patient outcomes based on input features.

Machine learning: A branch of artificial intelligence that enables computer systems to learn from data and improve performance on specific tasks without being explicitly programmed. Machine learning algorithms are widely used in health surveillance to analyze large datasets and extract valuable insights for public health interventions.

Model evaluation: The process of assessing the performance of machine learning models using various metrics such as accuracy, precision, recall, and F1 score. Model evaluation is essential in health surveillance to ensure that AI algorithms are reliable and effective in predicting health outcomes.

Natural language processing (NLP): A subfield of artificial intelligence that focuses on the interaction between computers and humans using natural language. NLP techniques are applied in health surveillance to extract valuable information from unstructured text data such as medical records and clinical notes.

Overfitting: A common issue in machine learning where a model learns the noise in the training data rather than the underlying patterns. Overfitting can lead to poor generalization and inaccurate predictions. Techniques such as regularization and cross-validation are used in health surveillance to prevent overfitting.

Precision medicine: An approach to healthcare that considers individual variability in genes, environment, and lifestyle to tailor medical treatments to the specific needs of patients. Machine learning techniques are integral to precision medicine in health surveillance for identifying biomarkers, predicting treatment responses, and optimizing patient outcomes.

Random forest: An ensemble learning technique that combines multiple decision trees to make predictions.

Random forests are popular in health surveillance for tasks such as disease classification, risk assessment, and feature importance ranking.

Reinforcement learning: A machine learning paradigm where an agent learns to make sequential decisions by interacting with an environment and receiving rewards or penalties. Reinforcement learning is applied in health surveillance for optimizing treatment strategies, resource allocation, and policy interventions.

Support vector machines (SVM): A type of supervised learning algorithm used for classification and regression tasks. SVMs are effective in health surveillance for tasks such as patient outcome prediction, disease diagnosis, and medical image analysis.

Time series analysis: A statistical technique used to analyze and forecast trends in data collected over time. Time series analysis is critical in health surveillance for monitoring disease outbreaks, predicting patient outcomes, and identifying temporal patterns in healthcare data.

Transfer learning: A machine learning technique that leverages knowledge from one domain to improve performance in another related domain. Transfer learning is useful in health surveillance for tasks such as disease diagnosis, patient risk prediction, and medical imaging analysis.

Unsupervised learning: A machine learning paradigm where models learn patterns and structures in data without labeled examples. Unsupervised learning algorithms are applied in health surveillance for tasks such as clustering, anomaly detection, and dimensionality reduction.