
Graduate Certificate in AI for Renewable Energy Forecasting

Machine Learning for Renewable Energy Data

Artificial Intelligence (AI): The simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving.

Machine Learning (ML): A subset of AI that provides systems the ability to automatically learn and improve from experience without being explicitly programmed. It focuses on the development of computer programs that can access data and use it to learn for themselves.

Renewable Energy (RE): Renewable energy is energy that is collected from resources that are naturally replenished on a human timescale, such as sunlight, wind, rain, tides, waves, and geothermal heat.

Renewable Energy Forecasting: The process of predicting the amount of renewable energy that will be available at a given time in the future. This can help grid operators to balance supply and demand, and ensure that there is enough energy to meet demand.

Supervised Learning: A type of machine learning where the model is trained on a labeled dataset. In other words, the correct answer (label) is provided for each example in the training data.

Unsupervised Learning: A type of machine learning where the model is trained on an unlabeled dataset. The model must find patterns and relationships in the data without any guidance.

Reinforcement Learning: A type of machine learning where an agent learns to make decisions by taking actions in an environment to maximize some notion of cumulative reward.

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, requires less data preprocessing by humans, and can often produce more accurate results than traditional machine learning approaches.

Time Series Forecasting: The use of a model to predict future values based on past values. In the context of renewable energy, this can be used to predict the amount of energy that will be generated by a wind turbine or solar panel in the future.

Feature Engineering: The process of selecting and transforming raw data features into ones that better represent the underlying problem to the predictive models, thereby improving their performance.

Data Preprocessing: The process of transforming raw data into an understandable format. This can include cleaning the data, normalizing the data, and handling missing values.

Overfitting: A modeling error that occurs when a model learns the detail and noise in the training data to the extent that it negatively impacts the performance of the model on new data.

Underfitting: A modeling error that occurs when a model is too simple to learn the underlying trend of the data.

Cross-Validation: A technique for assessing how the results of a statistical analysis will generalize to an independent data set. It is mainly used in backgrounds where the goal is prediction, and one wants to estimate how accurately a predictive model will perform in practice.

Regression Analysis: A statistical process for estimating the relationships among variables. It includes many techniques for modeling and analyzing several variables.

Classification Analysis: A statistical analysis technique used to predict the categorical class labels of new instances, based on past observations.

Evaluation Metrics: Quantitative measures used to evaluate the performance of a machine learning model. Common metrics include accuracy, precision, recall, and F1 score.

Bias-Variance Tradeoff: The problem of finding a balance between bias and variance such that the model is complex enough to fit the data well, but not so complex that it overfits the data.

Recurrent Neural Networks (RNNs): A class of artificial neural networks where connections between nodes form directed cycles. This creates an internal state of the network which allows it to exhibit dynamic temporal behavior.

Long Short-Term Memory (LSTM): A special kind of RNN that is able to remember information for long periods of time. It is often used in tasks that require the understanding of context, such as language translation and speech recognition.

Convolutional Neural Networks (CNNs): A class of deep, feed-forward artificial neural networks that have proven very effective in areas such as image recognition and classification.

Autoencoder: A type of artificial neural network used for learning efficient codings of input data. It is typically used for the purpose of dimensionality reduction or denoising.

Generative Adversarial Networks (GANs): A class of artificial neural networks used in unsupervised machine learning, implemented by a system of two neural networks contesting with each other in a zero-sum game framework.

Natural Language Processing (NLP): A field of AI 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.

Named Entity Recognition (NER): A subtask of NLP that seeks to locate and classify named entities in text into predefined categories such as person names, organizations, locations, medical codes, time expressions, quantities, monetary values, percentages, etc.

Part-of-Speech Tagging (POS Tagging): The process of marking up a word in a text as corresponding to a

particular part of speech, based on both its definition and its context.

Sentiment Analysis: The use of natural language processing, text analysis, and computational linguistics to identify and extract subjective information from source materials.

Topic Modeling: A type of statistical model for discovering the abstract "topics" that occur in a collection of documents.

Word Embedding: A technique in NLP where words or phrases from the vocabulary are mapped to vectors of real numbers. Conceptually it involves a mathematical embedding from a space with one dimension per word to a continuous vector space with a much lower dimension.

Transfer Learning: A research problem in machine learning that focuses on storing knowledge gained while solving one problem and applying it to a different but related problem.

Active Learning: A special case of machine learning where a learning algorithm can interactively query the user to obtain the desired outputs at new data points.

Federated Learning: A machine learning approach that allows for decentralized data analysis. Instead of bringing all data to a central server for training, federated learning brings the model to the data, allowing it to learn from data stored across multiple devices or servers.

Explainable AI (XAI): A subfield of AI focused on creating a suite of machine learning techniques that produce more explainable models while maintaining a high level of learning performance.

Interpretable Machine Learning: A subfield of AI focused on creating machine learning models that deliver more explainable predictions and are thus more understandable by human experts.

Feature Importance: A measure of how useful each feature is in making predictions with a machine learning model.

Shapley Additive Explanations (SHAP): A method for interpreting the output of a machine learning model by quantifying the importance of each feature in making a specific prediction.

Local Interpretable Model-Agnostic Explanations (LIME): A method for explaining the predictions of any machine learning classifier by learning an interpretable model locally around the prediction.

Anomaly Detection: The identification of unusual patterns that do not conform to expected behavior, called outliers. It has many applications in business, from intrusion detection (identifying strange patterns in network traffic that could signal a hack) to system health monitoring (spotting a malignant tumor in an MRI scan), and from fraud detection in credit card transactions to fault detection in operating environments.

Natural Language Generation (NLG): The use of artificial intelligence to produce written or spoken narrative from structured data.

Data Augmentation: A strategy that enables practitioners to significantly increase the diversity of data available for training models, without actually collecting new data.

Data Drift: The change in the distribution of data over time. It is a common challenge in real-world machine learning applications, where models are often trained on historical data and then deployed to make predictions on new, unseen data.

Concept Drift: The change in the underlying relationship between the input features and the output label over time.

Covariate Shift: A type of dataset shift