**CSE4288**

**Introduction to Machine Learning**

K-Nearest Neighbor Classifier

Mohamad Nael Ayoubi

150120997

* **Introduction**

The k-Nearest Neighbor (k-NN) Algorithm: The k-Nearest Neighbor (k-NN) algorithm is a simple, yet powerful supervised learning method widely used for classification and regression tasks. It is a lazy learner, meaning it does not build an explicit model during the training phase. Instead, it stores the training data and makes predictions based on the majority class of the k closest neighbors in feature space. All instances correspond to points in the n-D space. The closeness is determined by a distance metric, such as Euclidean or Manhattan distance.

The objective of this assignment is to understand and implement the k-NN algorithm from scratch without using machine learning libraries like scikit-learn. The tasks include:

1. Preparing the Play Tennis dataset.
2. Implementing the k-NN algorithm for classification.
3. Evaluating its performance using metrics such as accuracy and a confusion matrix.
4. Analyzing the results and identifying any challenges during implementation.

* **Methodology**

1. Data Preparation:

The Play Tennis dataset consists of categorical features (e.g., Outlook, Temperature, Humidity, Wind) and a target variable (PlayTennis), which were converted to numeric form using label encoding. As shown below: