

AIDI 1002: AI ALGORITHMS I

EXERCISE 1 ML ALGORITHMS

SAHIL KHAN

100942935

REGRESSION PROBLEMS

CUSTOMER CHURN PREDICTION

Customer Churn is a problem that is faced by many organizations like telecom companies or energy distributor organizations. It refers to the situation where customers stop using the services/products of their provider company and start using of the different one. Predicting customer churn plays an important role in the retention of the customers and identify beforehand how many customers or percentage of customers are planning or at risk of leaving, allowing for targeted retention strategies.

Objective: Predict the probability of a customer leaving the subscription service based on historical usage patterns, interactions, feedback, and other relevant features.

Target Variable: Probability of churn (a continuous value between 0 and 1).

Regression Aspect: The target variable is a continuous probability, making it a regression problem. The model learns from historical data to estimate the likelihood of a customer churning.

CUSTOMER LIFETIME VALUE

Customer Lifetime Value (CLV) is a measure of the anticipated revenue a firm may expect to receive from a client over the course of their entire business relationship. Making intelligent marketing and consumer acquisition decisions is aided by accurate CLV estimates.

Objective: Predict the monetary value a customer is expected to generate over their lifetime with the business based on past purchasing behavior and other relevant features.

Target Variable: Monetary value representing the customer's expected lifetime value (a continuous numerical value).

Regression Aspect: The target variable is a continuous numerical value representing the expected lifetime value, making it a regression problem. The model learns to predict a numerical value indicating the potential revenue generated by a customer over their lifetime.

ENERGY EFFICIENCY OPTIMIZATION

Optimization in Energy Efficiency is the solution which every company/organization is looking for. It deals with the prediction of energy efficiency of a system based on various dependent variables like weather conditions, equipment usage, architectural design and historical data of energy consumption in a industrial process or a company owned building.

Objective: Optimize energy consumption and reduce wastage by predicting and improving the energy efficiency of the system.

Target Variable: Energy efficiency score or a continuous numerical value representing the energy efficiency of the system.

Regression Aspect: It involves predicting a continuous numerical value (energy efficiency) based on input features to optimize and improve energy consumption in a system.

CLASSIFICATION PROBLEMS

SENTIMENT ANALYSIS

Sentiment Analysis is a classification machine learning problem in which the objective is to determine the sentiment of text data such as tweet or reviews whether it is in positive, negative or neutral from social posts, reviews or comments.

Objective: The objective is to classify the text into predefined sentiment categories to understand public opinion or customer feedback.

Target Variable: The target variable is categorical, indicating the sentiment category (e.g., positive, negative, neutral).

Classification Aspect: This is a multi-class classification problem where the algorithm learns to assign a specific sentiment label to each piece of text based on the language used and the emotions expressed.

INSURANCE FRAUD DETECTION

The problem in insurance fraud detection is to identify fraudulent insurance claims based on various features associated with each claim. Insurance fraud occurs when individuals or entities deceive insurance companies to receive undeserved payouts or benefits, leading to financial losses for the insurance providers.

Objective: The goal is to develop a system that can automatically and accurately identify potentially fraudulent claims, enabling timely investigation and action by the insurance company.

Target Variable: The target variable is binary, indicating whether a claim is fraudulent or non-fraudulent.

Classification Aspect: This problem falls under the domain of binary classification. Machine learning models are trained on historical insurance claim data, where each claim is labeled as either genuine or fraudulent based on investigations or follow-ups.

SPECIES CLASSIFICATION IN ECOLOGY

In the field of ecology, species classification involves categorizing different species of plants or animals based on a range of features and attributes, such as physical characteristics (e.g., leaf shape, flower color), habitat preferences, behavioral patterns, and genetic markers.

Objective: The main objective of this problem is to develop a reliable and accurate classification model that can automatically identify and categorize species based on their unique traits.

Target Variable: The target variable is categorical, representing the species or class to which an organism belongs (e.g., species A, species B, species C). Each category corresponds to a specific type of plant or animal species.

Classification Aspect: This is a multi-class classification problem, as there are multiple species to classify. Machine learning models are trained using labeled datasets, where each example is associated with a particular species label.

UNSUPERVISED ML PROBLEMS

CUSTOMER PREFERENCE ANALYSIS

The problem is to understand and analyze customer preferences and behavior based on their browsing history, interactions, and purchase patterns on an e-commerce platform.

Objective: The objective is to cluster customers into groups with similar preferences, aiding personalized product recommendations and marketing strategies.

Target Variable: There is no specific target variable; the algorithm clusters customers based on their behavior.

Unsupervised Aspect: This is an unsupervised learning problem because the algorithm identifies clusters of customers based on their behavior and preferences without relying on predefined labels.

FASHION APPAREL CLASSIFICATION

The problem involves classifying different fashion apparel items (e.g., shirts, pants, dresses) based on their visual characteristics and features captured through images.

Objective: The objective is to cluster the fashion apparel images into meaningful categories or clusters, without predefined labels or prior knowledge of the specific types of apparel. This clustering will help organize and categorize the apparel based on visual features, enabling better inventory management, recommendation systems, or e-commerce applications.

Target Variable: The algorithm will group similar apparel images into clusters, allowing us to define categories based on visual similarities.

Unsupervised Aspect: This problem is unsupervised because we are not provided with labeled training data specifying the type of each fashion apparel item. The algorithm must learn patterns and features from the images itself and cluster them based on similarities in visual characteristics.

DOCUMENT SUMMARIZATION

The problem is to generate concise and coherent summaries of documents, such as articles, research papers, or reports.

Objective: The objective is to extract the most important information and key points from the text, reducing the document's length while preserving its meaning.

Target Variable: There is no specific target variable; the algorithm generates summaries.

Unsupervised Aspect: This can be framed as an unsupervised learning problem, particularly in extractive summarization, where the algorithm identifies and selects important sentences or phrases based on the document's content and structure without using labeled data.