**INF 2179 – Winter 2025 Classification Challenge**

**Hamid Parsazadeh**

**Introduction**

The goal of this challenge is to develop predictive machine learning models using women’s clothing ecommerce review data. Participants will tackle two classification tasks:

1. Recommendation Prediction: Predict whether a user recommends a product (binary classification: recommended = 1, not recommended = 0).

2. Star Rating Prediction: Predict the user’s given star rating (1 to 5 stars) using the review text and title.

**Task 1: Recommendation Prediction**

In this task we need to create and train a model to predict customer’s opinion (recommended or not) on various products of a clothing company based on their age (18 to 99) and their review. The target variable has only two values: 0 or 1 and the features are mixed: numerical, textual or categorical variables.

We need to convert the both categorical and textual variables into numerical values so the predicting models can process. Therefore we use ***Label Encoder*** method for categorical variables and ***TF-IDF technique*** to transform texts in textual variables to numerical vectors