Smart Tariff & Fee Assistant (STAFF)

CMPT-310: Milestone 1 — Progress Report (Oct 26) — Group 26

Group Members: **Kazi Boni Amin (kba109@sfu.ca) #301591325, Ertugrul Yurtseven (eya38@sfu.ca) #301588905, Brayden Yee (@sfu.ca) #301559654**

1. Project Summary

GitHub Link: https://github.com/Marchosias405/STAFF.git

Our project aims to predict Harmonized System (HS) tariff codes, and shipping restrictions for e-commerce items using structured data. The system uses product tags, price, weight, and origin/destination as inputs and outputs the most suitable HS codes with compliance flags. We use models which include KNN and ID3 Decision Trees for classification.

2. Accomplishments So Far:

- Prepared cleaned datasets and rule files (samples_clean.csv, rules_*.csv).
- Created feature preprocessing pipeline (such as one-hot encoding).
- Implemented KNN and ID3 Decision Tree models with training/evaluation scripts and metrics.
- Built a CLI tool for single-item predictions.
- Added a rule engine to flag restricted or gift items.
- Completed EDA notebook and added clear documentation with run instructions.

3.Fall Backs / Challenges:

- Deferred BERT/text embeddings to keep the system reproducible and simple (after TA feedback).
- We still need to fix ID3 hyperparameters and expand rule coverage.
- Some CSV files required cleanup, and small data size affects test accuracy.

4. Project Changes:

- Switched from BERT to structured features based on the TA feedback to make the system more explainable, reproducible, and faster to train.
- Focused on interpretability and Top-k outputs (based on Accuracy) to provide more useful predictions for real-world cases.