Executive Summary: Course Review Analyzer

Written by Reggie Bain

In this project, we present a Natural Language Processing (NLP) pipeline for analyzing course reviews, with the goal of filtering out low-quality entries and extracting actionable sentiment insights. The work culminates in a Streamlit web application that allows users to input and assess course reviews in real time.

Objectives

To evaluate, clean, and analyze course reviews using modern NLP and machine learning techniques to:

- Classify review sentiment
- Detect and filter meaningless or gibberish reviews
- Summarize insights for course improvement and broader performance metrics

Stakeholders

The key beneficiaries of this work include:

- Instructors seeking direct feedback
- Academic and industry administrators needing scalable quality metrics
- Students and consumers making informed decisions based on curated reviews

Data Sources & Exploration

The project utilizes:

- A large dataset of Coursera course reviews (via Kaggle)
- Amazon product reviews with labeled gibberish entries (also via Kaggle)

Modeling Pipeline

The modeling was structured into three core phases:

1. **Entropy Analysis**
    - Investigated the entropy of reviews across different languages.
    - Found statistically significant differences using non-parametric tests (e.g., Kruskal-Wallis).
2. **Gibberish Detection**
    - Trained a classifier using Amazon data to detect meaningless reviews.
    - Applied statistical tests (f-statistics) and tree-based models to identify key features like entropy, word count, punctuation ratio, and language detectability.
    - Achieved substantial improvement over baseline methods in identifying low-quality text.
3. **Sentiment Analysis**
    - Leveraged both classical NLP features and fine-tuned deep learning models to assess sentiment.
    - Created features from token patterns, sentiment lexicons, and embeddings.
    - Demonstrated that sentiment could be predicted with better classification metrics over various relevant baseline models.

Deployment

- Deployed Streamlit tool that allows users to track sentiment, summarize, calculate an overall score, and get constructive feedback
- The app uses a simplified version of the models due to space/compute limitations on their free tier hosting but provides a proof-of-concept of the overarching idea of the project.